

Perspective Analysis for Online Debates

Sukru Tikves, Sedat Gokalp, Mhamed Temkit, Sujogya Banerjee, Jieping Ye, Hasan Davulcu
Arizona State University, P.O. Box 87-8809, Tempe, AZ, 85281 USA
{Sukru.Tikves, Sedat.Gokalp, Mhamed.Temkit, Sujogya.Banerjee, Jieping.Ye, Hasan.Davulcu}@asu.edu

Abstract—Internet and social media devices created a new public space for online debate on political and social topics. A debate is defined as a formal discussion on a set of related topics in a public meeting, in which opposing perspectives and arguments are put forward. In this paper, we develop automated perspective discovery techniques which would contribute to the understanding of features (i.e. social, political, cultural, religious beliefs, goals, and practices) shared by each side of the debate. Secondly, we show that, compared to a semi-automated process, our perspective discovery algorithms not only identify larger number of relevant features, but they also yield a higher accuracy scaling of moderate to extreme organizations on both sides of a debate.

I. INTRODUCTION

Internet and social media devices created a new public space for debate on political and social topics [1], [2]. Hotly debated issues span all spheres of human activity; from liberal vs. conservative politics, to radical vs. counter-radical religious debate, to climate change debate in scientific community. Many prominent 'camps' have emerged within Internet debate rhetoric and practice [3]. There are many applications [4]–[8] for recognizing politically-oriented sentiment in texts.

A debate is defined as a formal discussion on a set of related topics in a public meeting, in which opposing arguments are put forward. Initially, we observe that given a certain topic, each organization's web site mostly discusses their own perspectives related to that topic, and occasionally discusses others' perspectives, relating them back to their own perspectives. As a case study of an ongoing large scale online debate, we utilize the discourse found in the web sites of 10 *radical*, and 13 *counter-radical* Indonesian religious organizations - comprising a total of 37,000 articles dating from 2001 to 2011. Radicalism [9] is the ideological conviction that it is acceptable and in some cases obligatory to use violence to effect profound political, cultural and religious transformations and change the existing social order. Counter-radicals oppose violent social and political movements.

In our prior work [10] we showed that both counter-radical and radical movements in Muslim societies exhibit distinct combinations of perspectives on various social, political, and religious issues, and those perspectives can be mapped to a latent linear continuum, or a scale. The resulting model allowed us to measure the distance between organizations and movements over the underlying scale. It also facilitates tracking the ways in which movements and organizations change over time and space [11].

In [10], we utilized a simple term frequency - inverse document frequency (TF-IDF) [12] based technique to generate a

large candidate list of topics and perspectives for inclusion in scaling analysis. Top 100 n-grams from each organization's web site were collected into a list of candidate keywords. Next, we asked social scientists to scan this list manually, and identify all significant keywords belonging to social, political, economic, and religious perspectives. During this process, social scientists on our team assessed a total of 790 candidate keywords; of which 29 and 26 were selected by experts for inclusion in the radical and counter-radical scaling analysis respectively.

Upon analyzing the results of this study, we have identified that automatically generating the items of the radical and counter-radical scales would be an important contribution to the research. For example, among the included scale items were phrases like "religious education". However, reaching that item from a seed topic (like "education"), instead of manual selection would be desirable. This would not only decrease the expert intervention in scale generation, it would also provide us with useful perspective of organizations on these topics aligned with the underlying scale. In order to explore this idea, we have developed methods for perspective analysis built upon previous findings of the scaling research.

In this paper, our primary contribution is the development of automated perspective discovery techniques which would contribute to the understanding of features (i.e. social, political, cultural, religious beliefs, goals, and practices) shared by one side of a debate, and by those opposing them. Secondly, we show that, our perspective discovery algorithms not only identify larger number of relevant features - compared to the semi-automated process, but also yield a higher accuracy scale of radicalism vs. counter-radicalism.

II. PROBLEM DEFINITION

In this study, we would like to automatically discover the perspectives of organizations on a given set of topics, and extract the underlying discourse scale. Here, a scale is a social science model to measure the positioning of organizations on a latent dimension. We would like this generated scale to fit both the theoretical model, and also be able to mimic expert level deduction in the domain.

For our specific instance, we have utilized Rasch model [13] as the probabilistic version of a Guttman scale, and web mining techniques for producing response tables from organizations' corpora. We have utilized the opinion of our experts on Indonesian Islamic religious organizations as the target gold standard of orderings of these organizations.

III. METHODS AND THEORY

A. Overall System Model

We devised an end-to-end pipeline of methods that would generate two sets of perspectives for each of the polarities of a scale, starting from a mined web corpus, and a list of topics provided by experts.

Overall flow of the system consists of data gathering from experts, and web mining, manual topic selection, perspective analysis, and scale generation.

B. Scaling

In social science *scaling* is a process of measuring and ordering entities such as *subjects* based on their qualitative attributes called *items*. In general, subjects respond to surveys in form of interviews or questionnaires, where items are presented to the subjects in form of questions. Some of the widely followed scaling procedures in social science are Likert scale [14], Thurstone scale [15], and Guttman scale [16]. Guttman scaling procedure orders both the subjects and the items simultaneously with respect to some underlying latent cumulative continuum. In this study, items tend to have a natural total ordering to partial ordering, since an organization support for keyword such as “Sharia” will most likely imply their support for the keyword “Quran”, we used the Guttman scaling to rank the organizations based on their response on the radical and counter-radical keywords.

A Guttman [17] scale presents a number of items to which each subject is requested to provide a dichotomous response, e.g. agree/disagree, yes/no, or 1/0. This scaling procedure is based on the premise that the items have strict orders (i.e., the items are presented to the subjects ranked according to the level of the item’s difficulty). An item “A” is said to be “more difficult” than an item “B” if any subject answering “yes” on item “A” implies that the subject will also answer “yes” on item “B”. A subject who responds to an item positively is expected to respond positively to all the items of lesser difficulty.

Guttman scale is a deterministic process and the score of a subject depends on the number of affirmative responses he has made on the items. Scores in Guttman scale can also be interpreted as the “ability” of a subject in answering questions sorted in increasing order of “difficulty”. These scores when presented on an underlying scale, give us an ordering of the subjects based on their “ability” too.

The objective of our paper is to order the Indonesian Islamic organizations based on their views on religio-social keywords which have an inherent ordering. An organization supporting “Sharia” will also likely to believe in “Quran”. So it makes sense to use Guttman scaling procedure to rank the organizations and their beliefs and practices. One drawback of Guttman scale is that it is deterministic and assumes a strict ordering of the items. We used Rasch [18] model to overcome this drawback, by providing a probabilistic framework for Guttman scales. Specifically, in the simple Rasch model, the probability of a positive response (yes) is modeled as a logistic

function of the difference between the subject and item’s parameters. Item parameters pertain to the difficulty of items while subject parameters pertain to the ability of subjects who are assessed. A subject of higher ability relative to the difficulty of an item, has higher probability to respond to a question affirmatively. In this paper Rasch models are used to assess the organizations degree of being radical or counter-radical based on the religio-social keywords (items) appearing in their rhetoric.

Rasch model maps the responses of the subjects to the items in binary or dichotomous format, i.e., 1 or 0. Let Bernoulli variable X_{vi} denotes the response of a subject v to the item i , variable θ_v denotes the parameter of “ability” of the subject v and β_i denotes the parameter of “difficulty” of an item i . According to the Rasch model the probability that subject v responds 1 for item i is given by

$$P(X_{vi} = 1 | \theta_v, \beta_i) = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)}$$

The maximum likelihood method is used to provide estimates for subject and item parameters. We can also assess whether the data fits the model by looking at goodness of fit indices, such as the Andersen’s likelihood ratio test (LR-test) [19]. A p -value, returned by the test, indicates the goodness of fit and a p -value higher than 0.05 indicates no presence of lack of fit. We used the eRm [20] package to run the Rasch models

C. Implementing Rasch Model in the Text Mining Domain

Other works in text-mining domain such as sentiment analysis, have used Rasch model in their analysis [21]. In our application, Rasch model *subjects* correspond to a group of religious organizations, and *items* correspond to a set of keywords for socio-cultural, political, religious *radical* and *counter-radical* beliefs, and practices. An organization responding “yes” to a feature means the organization exhibits that feature in its narrative, while an organization responding “no” to a feature indicates that the organization does not exhibit such a feature. *Difficulty* of an item translates to *strength* of the corresponding attitude in defining radical or counter-radical ideology of any organization. Similarly *ability* of a subject in this case means the *degree* of radicalism or counter-radicalism exhibited by an organization’s rhetoric.

D. Our Initial Work on Scale Generation

Our initial work [10] depended on more direct interaction with experts’ opinion to build a model that can capture the underlying dynamics of the scale. The experts both provided a set of target organizations, and also directly selected the items that would make up the scale, from a machine generated candidate list. The candidate list consisted of the union of top-100 n-grams from each organization’s individual corpus, which were a total of 790 items. The resulting scale has utilized a total of 55 of keywords selected by experts.

E. Debates and Perspective Analysis

Upon inspecting the keywords selected by our team of experts we observed that, some of these keywords correspond to differing perspectives on a set of topics that are debated within these web sites. Definition of **debate** is “a formal discussion on a particular topic in a public meeting or legislative assembly, in which opposing arguments are put forward.”¹. During a debate on a particular topic, like education, both radical and counter-radical organizations discuss different perspectives – such as “secular multi-cultural education” vs. “sharia based religious education”.

During the design of an automated perspective detection algorithm, we made the following simplifying assumptions:

- 1) Organizations will *mostly* discuss their own perspective in a debate;
- 2) Organizations will *occasionally* mention others’ perspectives, however, then relate them back to their own perspective.

In our upcoming work [11], we present a mathematical formulation of the perspective keyword generation problem for a given topic, and provide an NP-Completeness proof of this problem, and design an exact solution through an ILP (integer linear programming) based solver.

The input to this algorithm also takes the polarity suggestion from experts into consideration, for automatically identifying the discriminating perspectives of those organizations from opposite sides of a debate.

However, due to the algorithmic complexity and the strict constraints of the exact model, the ILP based solver was not always able to produce acceptable solutions. Namely, for larger debated topics, the runtime requirements² exceeded acceptable limits of the study, and for more intervened debates, none of the possible item sets could satisfy strict constraints of the ILP definition.

In order to resolve this, in our current version of the system, we have worked with a feature selection framework, SLEP. The discussion of the implementation of SLEP is discussed in the next section.

F. SLEP: A Sparse Learning Package

In order to address the scalability problem encountered in ILP we resorted to SLEP [22], again with the underlined motivation to select a subset of discriminating features that can (a) classify and (b) satisfy Guttman scale [16]. The following steps describe our algorithm:

- 1) For each topic, calculate the frequency of the words occurring within a fixed size window of the topic keyword
- 2) Filter the term \times document matrix to include only the most frequent 1000 words from each camp
- 3) Formulate the problem in a general sparse learning frame [22]. Logistic formulation fits our application, since it is a dichotomous classification problem

¹Oxford Online Dictionary

²Given data volume projections, we have estimated an upper bound of one hour runtime restriction per topic. For this paper, we have run the cplex ILP solver several hours for each topic before a timeout.

$$\min_x \sum_{i=1}^m w_i \log(1 + \exp(-y_i(x^T a_i + c))) + \lambda \|x\|_1 \quad (1)$$

$$+ \frac{\rho}{2} \|x\|_2^2 \quad (2)$$

where D_i is the *document* i and F_j is the feature (*word*) j . A is the term \times document matrix with all $A_{ij} \geq 0$, $y_i \in y$ is the class of each document D_i coded as +1 for *Radical* (R) and -1 for *Counter-Radical* (CR) and x_j is the weight for each feature F_j . Let us explain further the three terms involved in the convex optimization problem.

- $\sum_{i=1}^m w_i \log(1 + \exp(-y_i(x^T a_i + c)))$, this first term is related to the logistic classification error. We set the weights w_i values to be all 1 so that all documents have the same weight.
- $\lambda \|x\|_1$, this term involving the L_1 norm deals with the sparsity of the solution vector x . We experienced with several lambda values which resulted with an x vector of various sparsity.
- $\frac{\rho}{2} \|x\|_2^2$, this last term deals with the ridge regression, which is an extra level of shrinkage. We set the weight of this term $\rho = 0$ as we were mainly driven by sparsity.
- We used the MATLAB implementation of the SLEP package³ which utilizes gradient descent approach to solve the aforementioned optimization problem. This package can handle matrices of 20M entries within a couple of seconds on a machine with standard configuration.
- The features with non-zero values on the x vector are the candidate discriminants. Let \mathcal{F}_R , where $x_j > 0$ be the discriminant for the R class. Similarly, let \mathcal{F}_{CR} , where $x_j < 0$ be the discriminant for the CR class due to the coding schema in step 3. Given that the optimized formulation resulted with a sparse x vector, most of the words F_j had $x_j = 0$ and hence were not included in either \mathcal{F}_R or \mathcal{F}_{CR} .

Note that the sets of features \mathcal{F}_R and \mathcal{F}_{CR} may not satisfy the Guttman pattern. These sets needed to be further filtered such that $\mathcal{F}'_R \subseteq \mathcal{F}_R$ and $\mathcal{F}'_{CR} \subseteq \mathcal{F}_{CR}$ would satisfy the Guttman pattern.

IV. SYSTEM ARCHITECTURE

A. Data gathering

Initially, social scientists were invited to use their domain and area expertise to identify a set of *organizations*, and hypothesize any number of unipolar or bipolar *scales* that could explain the variance among their beliefs and practices. Next, a set of web crawling scripts were created for extraction of articles from those organizations’ web sites. For each organization’s corpus we extracted their original text in Indonesian, and stored the corpus in a ZIP file.

B. Feature Extraction

After identifying the features for the analysis, we iterate over the documents in the corpus of each organization for the matching items. This yields a feature-document matrix.

³<http://http://www.public.asu.edu/~eje02/Software/SLEP>

This feature extraction task was performed in a simple three step procedure; initially the occurrence frequencies of particular features were counted within each organization’s corpus, then a threshold matrix was calculated from these initial values, and finally a binary response matrix was generated by applying these thresholds to the initial values.

The frequency metric is shown in formula 3, where k is the keyword, o is the organization, and D_o is the document set pertaining to that particular organization.

$$f_{o,k} = \frac{|\{d \mid k \in d, d \in D_o\}|}{|D_o|} \quad (3)$$

A threshold value for each keyword is calculated from the values in the related column. And then, each element was converted into a binary value by comparing it to the column’s threshold.

C. Model Fitting

We fit the Rasch model on two datasets - (1) radical organizations with radical keywords and (2) counter-radical organizations with counter-radical keywords. We used the eRm package in R, an open source statistical software package⁴, to fit a Rasch model to the dataset, and obtain the organizations’ scores on the latent scale, which are the subject parameter estimates (θ_v) discussed in previous section. The eRm package⁵ fits Rasch models and provides subjects or organizations parameter estimates.

D. Feature Expansion Algorithm

We have observed that including all of the newly discovered features in the scale resulted a poor performance. This is because, they neither provided the desired Guttman pattern, nor the resulting scale aligned with the expert opinion. However, exhaustively enumerating all possible subsets to find an optimal one would also be undesirable due to time complexity. Thus we have devised a greedy expansion based algorithm to select the items that make up the scale. It chooses a sufficiently optimal subset of these features by expanding an initial set, incrementally adding features that offer a higher performance.

One possible implementation is shown in the algorithm in Fig. 1. This greedy algorithm will start from an initial set of features I , and iteratively select the features that increase the performance of the solution. The performance of a solution is evaluated by the SOLVE function, which takes a candidate input, and returns the performance value according to expert agreement.

Each iteration of the loop (lines 3 – 15) tries to iteratively expand the current set of selected features (lines 12 – 14). First, it evaluates the performance of the currently selected subset (line 4), and then identifies each not yet selected feature that provides a performance increase (lines 7 – 8), and finally collects them into the selected feature set for the next iteration (lines 5, and 8 – 10). When it can no longer include any new features, the algorithm will terminate.

⁴<http://cran.r-project.org/>

⁵<http://r-forge.r-project.org/projects/erm/>

Fig. 1. Feature set expansion algorithm

```

1: procedure GREEDY-SELECTION( $I, C, \triangleright$ )
   initial features  $I$ 
   candidate features  $C$ 
   comparison function  $\triangleright$ 

2:    $S \leftarrow I$ 
3:   repeat
4:      $m \leftarrow \text{SOLVE}(S)$ 
5:      $N \leftarrow \emptyset$ 
6:     for all  $c \in C \setminus S$  do
7:        $p \leftarrow \text{SOLVE}(S \cup \{c\})$ 
8:       if  $p \triangleright m$  then
9:          $N \leftarrow N \cup \{c\}$ 
10:      end if
11:    end for
12:    if  $N \neq \emptyset$  then
13:       $S \leftarrow S \cup N$ 
14:    end if
15:  until  $N = \emptyset$ 
16:  return  $S$ 
17: end procedure

```

Fig. 2. Feature set expansion algorithm modification, enabling special handling of the empty initial set of features

```

1: for all  $c \in C$  do
2:    $p \leftarrow \text{SOLVE}(S \cup \{c\})$ 
3:   if  $p \triangleright m$  then
4:     if  $S = \emptyset \wedge p - m < \delta$  then
5:        $N \leftarrow \{c\}$ 
6:        $m \leftarrow p$ 
7:     else
8:        $N \leftarrow N \cup \{c\}$ 
9:     end if
10:  end if
11: end for

```

Another performance trade-off was done using the natural grouping of the features. Since the features in our problem are grouped by topic, we decided to keep these natural groupings, thus making each c in set C a collection of features.

Here the \triangleright comparison function will assume greater-than-or-equal-to semantics. This is because, while we want to have the best possible scoring features as possible, we also want to be able to have a larger set of perspectives that can be used to explain the underlying latent scale.

In order to be able to handle the case of an empty initial feature set (I), we expanded the algorithm as shown in Fig. 2. This modification (to lines 6 – 11 in the original algorithm) will choose the best available features in the first iteration that are within a score difference of δ of each other. This also assumes SOLVE function will return a sensible upper limit value when an empty set is given as its input.

Fig. 3. Feature elimination algorithm provided by the eRm package

```

1: procedure STEPWISEIT( $m$ , EVAL)
   Rasch Model (RM)  $m$ 
   Evaluation Function EVAL
2:    $r \leftarrow m$ 
3:   repeat
4:      $e \leftarrow \text{EVAL}(r)$ 
5:     if not FITS( $e$ ) then
6:        $i \leftarrow \text{LOWESTRANKEDFEATURE}(e)$ 
7:        $x \leftarrow r\$x \setminus \{i\}$ 
8:        $r \leftarrow \text{RM}(x)$ 
9:     end if
10:  until Max # of Steps, or FITS( $e$ )
11:  return  $r$ 
12: end procedure

```

E. eRm Iterative Item Elimination Algorithm

While the eRm Rasch analysis package already does trivial eliminations in the model (for example, ignoring full/empty 1/0 responses), it also provides an algorithm to clean up a model from features that do not adhere to the Rasch model/Guttman pattern.

The overall idea of the algorithm is summarized in the algorithm in Fig. 3. While our greedy feature selection algorithm worked by expanding a set of features, this algorithm works by going the opposite direction, and reducing the feature set in each step. Here $\$$ is the R member access operator, where $r\$x$ is the feature set of rasch model object r , and \setminus is set difference. Functions EVAL, and LOWESTRANKEDFEATURE are references to eRm provided facilities to evaluate, and find the worst contributing item of Rasch models.

V. EXPERIMENTAL EVALUATION

In order to measure the relation of the generated perspectives to the underlying scale, we have performed a series of experiments designed to compare their scaling capabilities to the gold standard ordering done by the experts.

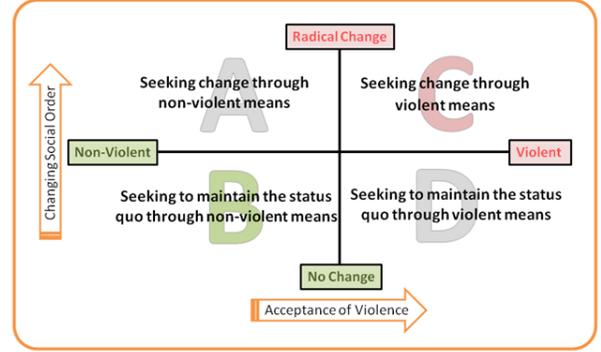
A. Indonesian Corpus

The corpus domain is the online articles published by the web sites of the 23 religious organizations identified in Indonesia, in the Indonesian language. These sources are the web sites or blogs of the identified think tanks and organizations. As discussed in the introduction, each source was classified as either radical or counter-radical by the area experts. We downloaded a total of 37,000 Indonesian articles published in these 23 web sites, dating from 2001 to 2011.

B. The Quadrants Model

Our project leverages the results of our previous work, which relied on social theory including Durkheim’s research on collective representations [23], Simmel’s work on conflict and social differentiation [24], Wallace’s writings on revitalization movements [25], and Tilly and Bayat’s studies on

Fig. 4. The quadrants model



contemporary social movement theory [26] [27]. Our team has also developed, and is currently testing a theoretically based class model comprised of continuous latent scales. The first pair of scales focus on distinctions between the goals and methods of counter-radical and radical discourse, and capture the degree to which individuals, groups, and behaviors aim to influence the social order (Change Orientation) and the methods by which they attempt to do so (Change Strategies).

Quadrants model (see Figure 4) captures multiple social trends in four quadrants A, B, C, and D, and it makes the significant distinction between violent and not-violent dimensions of both radicalisms and counter radicalisms. Using the quadrants model, a researcher can locate organizations, individuals, and discourses in broader categories while still considering subtle differences between groups within categories. A researcher can document movement and trends from category to category, and identify points where movement is likely.

C. Expert Opinion and Gold Standard of Rankings

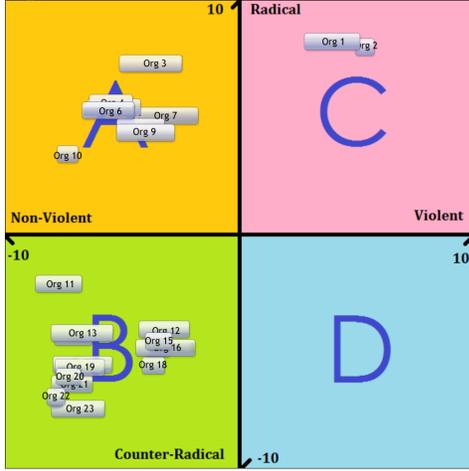
We collaborated with three area experts, who collectively possess 35 years of scholarly expertise on Indonesia and Islam. We utilized a homegrown graphical drag-and-drop user interface to collect their opinion to build the gold standard of the rankings. A screenshot of this tool is shown in Figure 5.

Each expert separately evaluated and ranked the organizations in the dataset according to a two dimensional scale of radical / counter-radical (R/CR) and violent / non-violent (V/NV) axis. The consensus among the experts was high; since per item standard deviations among the experts’ scores along the R/CR axis over a range of $[-10, 10]$, across all organizations were 2.75. The individual scores for each organization were combined and averaged to obtain the consensus *gold standard rankings* along the hypothesized R/CR scale.

In this paper, we used two measures for evaluating the difference between two separate rankings, based on Spearman’s footrule and Spearman’s correlation coefficient. The original work utilized a mean displacement based measure as follows:

Given two discrete ordering functions G , and R , on the organization set O , the normalized displacement of a single organization is given as:

Fig. 5. The visual interface of the expert opinion collector for manually placing the organizations on the two dimensional scale



$$disp(G, R, O, o) = \frac{|G(o) - R(o)|}{|O|} \quad (4)$$

Here, O is the set of organizations, G and R are one to one mapping functions of rankings from set O to range $[1, |O|]$. Then overall *error* measure for a given set of rankings was then defined as:

$$error(G, R, O) = \sum_{o \in O} \frac{disp(G, R, O, o)}{|O|} \quad (5)$$

In addition to this measure, we have also opted to include another measurement to take stability of the items into consideration. Based on the $L^2 - Norm$ of the normalized displacement function, the *msd* measure can be defined as the following:

$$msd(G, R, O) = \sum_{o \in O} \frac{disp(G, R, O, o)^2}{|O|} \quad (6)$$

Since our initial work, we have also modified the evaluation of the missing items. Specifically, for empty/full response patterns, the Rasch model would not be able to make any inference. Since we experimented with dynamic features, and the missing items varied in each test, we have opted to position them in their neutral places. This change has introduced a slight difference from the experimental results of our original study.

D. Baseline Performance

In our previous study [10] we have automatically generated a Rasch model from the organizational corpus data, and the expert selected items. We have observed that, against several baseline algorithms, including score sorting, and principal component analysis, the Rasch model was able to demonstrate the best available performance, and was ranked at expert level.

In order to have a baseline for comparison of the automatically generated items, we have opted to use this scale in our current study.

E. Candidate Perspectives

We have run both the ILP, and the SLEP based feature generators on all the 50 topics that has been identified. ILP was able to identify perspectives for 18 of the topics, while failed for the rest, due to either finding no viable exact solution, or timeouts. This resulted in a total of 2869 perspectives, with 159 average on each topic. Since these exact features also included items with very low support, we have filtered these results to include only the ones with higher frequency in the corpus. The final set contained a total of 227 perspectives on all 18 solved topics. On the other hand, SLEP was able to successfully generate candidate perspective on every 50 topic, totaling 1065 perspectives, with an average of 21 on each topic.

F. Aligning Perspectives with the Scales

In order to identify the perspectives that make up the theoretical scale we are working on (R/CR bi-polar scales on Indonesian Islamic religious organizations), we have devised a set of experiments that measure their relation to the Rasch model, and the expectation of field experts.

Initially, as a baseline, we have re-run the original scale with the expert selected features, with the new evaluation metrics. The mean displacement of the features was 0.1172, while the mean square displacement score was 0.0287. (The slight difference with the original paper is due to the handling of the missing items, discussed in Section V-C).

In order to observe the effect of the STEPWISEIT, we have run the elimination algorithm on the original set of features. The mean displacement was decreased to 0.1115, while the mean square displacement stayed the same. The algorithm has eliminated 15 features to reach this score. The summary of these experiments can be seen in Table I

TABLE I
EXPERIMENTAL RESULTS FOR THE ORIGINAL EXPERT SELECTED FEATURE BASED SCALES

	error	msd	run time
Original	0.1172	0.0287	14s
Original + STEPWISEIT	0.1115	0.0287	53s

G. The Initial Experiments with Feature Expansion

After establishing the baseline, we evaluated the perspective based features discovered by the ILP solver. First we built a model including all the candidate features proposed by the solver. This resulted in an mean displacement of 0.1323 and mean square displacement of 0.0284. While the performance was near the expert level, the hand selected features performed (13%) better than this initial run.

Then the features were refined with STEPWISEIT, and our GREEDY-SELECTION algorithms. The STEPWISEIT failed to provide better results, and actually performed worse, with mean displacement of 0.1632, and mean square displacement of 0.0386, while failing the *LR-test* for Rasch model fitness. The likely reason for this is that STEPWISEIT performs item

TABLE II

THE TOPICS CHOSEN BE THE GREEDY-SELECTION ALGORITHM FROM THE CANDIDATE PERSPECTIVES OF THE ILP SOLUTION.

Iteration	Topics
1	kufur disbelief
2	kdr, ke kafiran, kesetaraan, konstitusional, multikultural, sekularisme, tabligh, toleransi (domestic violence, infidelity, equality, constitutional, multicultural, secularism, tabligh, tolerance)
3	bunuh, gender, homoseksual, musyrikin, syirik (suicide, gender, homosexuals, idolaters, paganism)

eliminated locally based on individual item fitness, but the sparse nature causes loss of global Guttman pattern.

When we built an optimum item set from scratch using the GREEDY-SELECTION algorithm, we were able to identify 14 topics that contributed with better fitting perspectives. The expanding topic sets can be seen in Table II. The final solution had a mean displacement of 0.1020, with a mean square displacement of 0.0189. An additional cleanup using the STEPWISEIT algorithm over this existing solution did not produce better results.

The summary of these experiments can be seen in Table III.

TABLE III

SCALING EXPERIMENTS WITH THE ILP SOLVER BASED DATA

	error	msd	run time
ILP	0.1323	0.0284	3m:38s
ILP + StepWiseIt	0.1632	0.0386	56m:47s
Greedy(ILP)	0.1020	0.0182	
Greedy(ILP) + StepWiseIt	0.1122	0.189	

H. SLEP Based Features

In addition to the ILP based exact features, we also ran separate experiments for the SLEP output. These yielded a total of 449 features on counter radical, and 616 features on the radical scales. The overall runtime duration was 6 hours and 4 minutes. The resulting scales had a mean displacement of 0.1398 and mean square displacement of 0.0312. We opted not to run the STEPWISEIT on this particular case, since the expected runtime would be in the order of weeks, which would not be practical for the real life conditions of the project.

TABLE IV

SCALING EXPERIMENTS WITH THE SLEP SOLVER BASED DATA

	error	msd	run time
SLEP	0.1398	0.0312	6h:04m
Greedy(SLEP)	0.0982	0.0189	

Like the ILP based candidates, we also ran the GREEDY-SELECTION algorithm on the SLEP input (Table IV). Over two iterations, the algorithm was able to identify 15 topics, whose perspectives were closely related to the underlying scale. The expanding topic set can be seen in Table V. The best mean displacement achieved was 0.0982, with a corresponding mean square displacement of 0.0189.

Fig. 6. Runtime performance of the Rasch model fitting algorithm in the eRm package. The x axis corresponds to the number of items, while the y axis represents the runtime length in seconds. Notice that the scatter plot shows fitness to the x^2 polynomial prediction line.

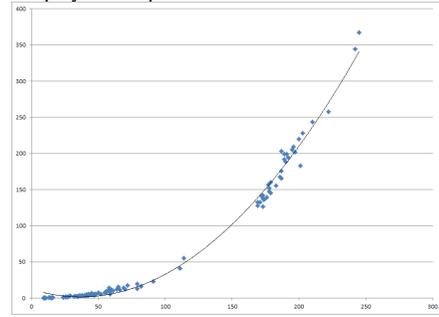


TABLE V

THE TOPICS CHOSEN BE THE GREEDY-SELECTION ALGORITHM FROM THE CANDIDATE PERSPECTIVES OF THE SLEP SOLUTION.

Iteration	Topics
1	manusia (human)
2	beragama, bunuh, dakwah, demokrasi, jihad, kafir, kristen, liberal, multikultural, pluralisme, politik, sipil, syariat, syirik (religion, kill, propaganda, democracy, jihad, infidel, Christian, liberal, multicultural, pluralism, political, civil, Sharia, polytheism)

The main reason that this table does not share a significant amount of topics with the ILP based topic set, is that the ILP solver could not provide results for the great majority of the topics selected by SLEP. The common ones, like “multikultural”, “syirik” were selected in both, while similar topics (like “politik”/“konstitusional”) were chosen when available.

I. Sample perspectives

A set of sample perspectives selected by the ILP solver are displayed in Figure 7. Here the columns represent individual topics, while two rows correspond to radical, and counter-radical perspectives on these topics. The items have been machine translated from Indonesian into English.

VI. CONCLUSION

In this paper, we have implemented a perspective identification algorithm, with exact, and approximate solutions, and also tested the viability of this algorithm, by trying to apply the discovered perspectives on an underlying bi-polar social scale.

In relation to our previous work, we have increased the automation in the scale generation process, by abstracting the hand selected items to automatically discovered perspectives from topics, and also increased the overall efficiency of the system, by producing lower distance to the expert agreement.

The theorized greedy growth based algorithm demonstrated the best available performance in our experiments. Additionally both ILP, and SLEP based perspective generation techniques provided features that fit the underlying bi-polar scale.

Fig. 7. A sample set of perspectives generated by the ILP based solver. Here each row represents a debate topic, while the linear scales represent the locations of the perspectives. The left side items are the counter-radical, and the right side items are the radical perspectives in each of these topics.

$$\min_x \sum_{i=1}^m w_i \log(1 + \exp(-y_i(x^T a_i + c))) + \lambda \|x\|_1$$

Our future work includes reducing the expert interaction further by automatically discovering debate topics, and investigating a possible use of these perspective analysis techniques in the related field of sentiment analysis.

REFERENCES

- [1] Z. Papacharissi, "The virtual sphere: the internet as a public sphere," *New Media Society*, vol. 4, no. 1, pp. 9–27, Mar. 2002. [Online]. Available: <http://nms.sagepub.com/cgi/content/abstract/4/1/9>
- [2] I. Himelboim, "Civil society and online political discourse: The network structure of unrestricted discussions," *Communication Research*, Oct. 2010. [Online]. Available: <http://dx.doi.org/10.1177/0093650210384853>
- [3] L. Dahlberg, "The internet and democratic discourse: Exploring the prospects of online deliberative forums extending the public sphere," pp. 615–633, Dec. [Online]. Available: <http://www.ingentaconnect.com/content/routledg/trics/2001/00000004/00000004/art00007>
- [4] T. Mullen and R. Malouf, "A preliminary investigation into sentiment analysis of informal political discourse," in *AAAI symp. on computational approaches to analysing weblogs (AAAI-CAAW)*, 2006, pp. 159–162.
- [5] R. Malouf and T. Mullen, *Graph-based user classification for informal online political discourse*, 2007.
- [6] M. Thomas, B. Pang, and L. Lee, "Get out the vote: Determining support or opposition from congressional floor-debate transcripts," in *In Proceedings of EMNLP*, 2006, pp. 327–335.
- [7] M. Bansal, C. Cardie, and L. Lee, "The power of negative thinking: Exploiting label disagreement in the min-cut classification framework," *Proceedings of COLING: Companion volume: Posters*, pp. 13–16, 2008.
- [8] W. Lin and A. Hauptmann, "Are these documents written from different perspectives?: a test of different perspectives based on statistical distribution divergence," in *Proc. of the 21st Int. Conf. on Computational Linguistics and the 44th annual meeting of the ACL*. ACL, 2006, pp. 1057–1064.
- [9] M. Woodward, I. Rohmaniyah, A. Amin, and D. Coleman, "Muslim education, celebrating islam and having fun as counter-radicalization strategies in indonesia," *Perspectives on Terrorism*, vol. 4, no. 4, 2010.
- [10] S. Tikves, S. Banerjee, H. Temkit, S. Gokalp, H. Davulcu, A. Sen, S. Corman, M. Woodward, I. Rochmaniyah, and A. Amin, "A system for ranking organizations using social scale analysis," in *EISIC*. IEEE, 2011, pp. 308–313.
- [11] S. Tikves, S. Banerjee, H. Temkit, S. Gokalp, H. Davulcu, A. Sen, S. Corman, M. Woodward, S. Nair, I. Rohmaniyah, and A. Amin, "A system for ranking organizations using social scale analysis," *Social Network Analysis and Mining (SNAM)*, Accepted for publication.
- [12] J. Hartigan and M. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.
- [13] G. Rasch, "On general laws and the meaning of measurement in psychology," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Psychology*, 4, 1961, p. 332.
- [14] R. Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, vol. 140, pp. 1–55, 1932.
- [15] L. L. Thurstone, "Attitudes can be measured," *American Journal of Sociology*, vol. 33, pp. 529–554, 1928.
- [16] J. McIver and E. Carmines, *Unidimensional Scaling*. Sage Publications, Inc, 1981, vol. 24.
- [17] L. Guttman, "The basis for scalogram analysis," *Measurement and prediction*, vol. 4, pp. 60–90, 1950.
- [18] D. Andrich, *Rasch models for measurement*. Sage, 1988.
- [19] D. Hessen, "Likelihood ratio tests for special rasch models," *Journal of Educational and Behavioral Statistics*, vol. 35, no. 6, p. 611, 2010.
- [20] P. Mair and R. Hatzinger, "Extended rasch modeling: The erm package for the application of irt models in r," 2007.
- [21] D. Drehmer, J. Belohlav, and R. Coye, "An exploration of employee participation using a scaling approach," *Group & Organization Management*, vol. 25, no. 4, p. 397, 2000.
- [22] J. Liu, J. Chen, and J. Ye, "Large-scale sparse logistic regression," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 547–556.
- [23] E. Durkheim, "The cultural logic of collective representations," *Social theory the multicultural and classic readings*, Wesleyan University: Westview Press, 2004.
- [24] G. Simmel, *Sociological Theory*. New York: McGraw-Hill, 2008.
- [25] A. Wallace, "Revitalization movements," *American Anthropologist*, vol. 58, pp. 264–281, 1956.
- [26] C. Tilly, *Social Movements*. Boulder, CO, USA: Paradigm Publishers, 2004.
- [27] A. Bayat, *Making Islam Democratic: Social Movements and the Post-Islamist Turn*. Stanford University Press, 2007.