

A System for Ranking Organizations using Social Scale Analysis

Sukru Tikves, Sujogya Banerjee, Hamy Temkit, Sedat Gokalp, Hasan Davulcu, Arunaba Sen, Steven Corman, Mark Woodward
Arizona State University, P.O. Box 87-8809, Tempe, AZ, 85281 USA
{stikves, sujogya, mtemkit, sgokalp, hdavulcu, asen, scorman, mataram}@asu.edu
Tel:(480) 965-6385, Fax: (480) 965-2751

Inayah Rochmaniyah, Ali Amin
Center for Religious and Cross Cultural Studies
Gadjah Mada University, Yogyakarta, Indonesia
{rochmaniyah, aleejtr77}@yahoo.com

Abstract—In this paper we utilize feature extraction and model fitting techniques to process the rhetoric found in the web sites of 23 Indonesian religious organizations – comprising a total of 37,000 articles dating from 2005 to 2011 – to profile their ideology and activity patterns along a hypothesized radical/counter-radical scale. We rank these organizations by assigning them to probable positions on the scale. We show that the developed Rasch model fits the data using Andersen’s LR-test. We create a gold standard of the ranking of these organizations through an expertise elicitation tool. We compute expert-to-expert agreements, and we present experimental results comparing the performance of three different baseline methods to show that the Rasch model not only outperforms our baseline methods, but it is also the only system that performs at expert-level accuracy.

I. INTRODUCTION

Being able to assess information on radical and moderate actors in a geographic area is an important research topic for our national security. *Radicalism* 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 fundamentally. Muslim radical movements have complex origins and depend on diverse factors that enable translation of their radical ideology into social, political and religious movements. In [1] Crelinsten states that “both violence and terrorism possess a logic and grammar that must be understood if we are to prevent or control them”. Therefore, analysis of Muslim radical and counter-radical movements requires attention to the global, national and local social, economic and political contexts in which they are located. Similarly, in the Islamic context, counter-radical discourse takes various different forms; discursive and narrative refutations of extremist claims, symbolic action such as ritual and other religious and cultural practices, and Islamic arguments for pluralism, peaceful relations with non-Muslims, democracy, etc. The most effective counter-radicals are likely to be religiously conservative Muslims. Effective containment and defeat of radicalism depends on our ability to recognize various levels of radicalization, and detection of counter-radical voices.

In our previous work [2], we attempted a clustering approach to obtain “natural groupings” of a number of local non-government religious social movements and organizations in Indonesia. Social scientists on our team observed that clustering was not fully able to separate all counter-radical or

radical organizations into pure clusters. Pure radical clusters were easily identified due to high similarity among their support for violent practices. Pure counter-radical clusters were identified due to their strong reactionary opposition to violent practices through protests and rhetoric. But the rest of the groupings were mixed. We realized that binary labeling as counter-radical or radical does not capture the overlap, movement and interactivity among these organizations. In this paper we hypothesize that both counter-radical and radical movements in Muslim societies exhibit distinct combinations of discrete states comprising various social, political, and religious beliefs, attitudes and practices, that can be mapped to a latent linear continuum or a scale. Using such a scale, an analyst can determine where exactly along the spectrum any particular group lies, and also potentially where it is heading with its rhetoric and activity.

Given the complex nature of the task, such as regional differences in local cultures, beliefs and practices, and in the absence of readily available high accuracy parsers, highly structured religio-social ontologies, and information extraction systems; we decided to devise a multi-lingual non-linguistic text processing pipeline that relies on only statistical modeling of keyword frequency and co-occurrence information.

We worked with social scientists on our team to come up with an orthogonal model comprising of two primary dimensions. Both dimensions, (i) radical/counter-radical and (ii) violent/non-violent, are characterized as latent, partial orders of discrete beliefs and practices based on a generalization of item order in Guttman scaling [3] using a Rasch model [4]. A true Guttman scale is a deterministic process, i.e. if a social movement subscribes to a certain belief or practice, than it must also agree with all lower order practices and beliefs on the scale. Of course, such perfect order is rare in the social world. The Rasch model provides a probabilistic framework for Guttman scales to accommodate for incomplete observations and measurement errors.

In this paper we present feature extraction and model fitting techniques to process the rhetoric found in the web sites of 23 religious Indonesian organizations – comprising a total of 37,000 articles dating from 2005 to 2011. We aim to profile their ideology and activity patterns along the hypothesized radical/counter-radical scale, and rank them by assigning them

to probable positions on this scale [5]. We used the eRm¹ package to fit the Rasch model on this data set, and identify organizations' positions based on maximum likelihood estimation [6]. The automated ordering of organizations is formed by ranking the organizations according to their estimates on the latent scale. We show that the model fits the data using the Andersen's LR-test[7]. We also created a gold standard of the ranking of these organizations through an expert opinion elicitation tool, and through the opinions of three ethnographers on our team who collectively possess 35 years of scholarly expertise on Indonesia and Islam. We computed expert-to-gold standard agreements, as well as compared the performance of three different baseline computational methods to show that the Rasch model presented here not only performs the best among the baseline methods, but that it also performs at an expert level of accuracy.

A. Organization of the paper

Next section provides an introduction to the theory of Guttman Scaling and Rasch Models. Section III defines the problem, presents the system architecture, and the methods used to solve the problem. Section IV describes the Indonesian corpus, expert opinion elicitation tool, baseline computational methods, and experimental evaluations. Section V concludes the paper.

II. INTRODUCTION OF GUTTMAN SCALING AND RASCH MODEL

In social science *scaling* is a process of measuring and ordering entities called *subjects* based on their qualitative attributes called *items*. In general, subjects are requested to respond to surveys conducted by means of structured interviews or questionnaires. Items are presented to the subjects in form of questions. Statistical analysis of the response of the subjects on the questions about items are used in *scaling* the subjects. Some of the widely followed scaling procedure in social science surveys are Likert scale [8], Thurnstone scale [9], and Guttman scale [10]. Guttman scaling procedure orders both the subjects and the items simultaneously with respect to some underlying cumulative continuum. In this paper we follow the Guttman scaling process to rank the organizations based on their response on the radical and counter-radical keywords.

A. Guttman Scaling

A Guttman [3] 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. For example, in order to find out how extreme a subject's view is on Guttman scale, the subject is presented with the following series of items in question form: (1) Are you willing to permit immigrants to live in your country? (2) Are you willing to permit immigrants to live in your community? (3) Are you willing to permit immigrants to live in your neighborhood? (4) Are you willing to permit immigrants to live to your next door? and (5) Are you willing to permit your child to marry an immigrant? If the items form a Guttman scale, any subject agreeing with any item in this series will also agree with other items of lower rank-order in this series. 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. So, a score of 2 for a subject in the above Guttman scale not only means he has given affirmative response to two of the questions or items, but also indicates that he agrees with two particular questions, namely the first and second. 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. For example, two such keywords are "Quran" and "Sharia". 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. In real world, it is difficult to order all the items in such a strict level of increasing difficulty, therefore perfect scales are not often observed in practice. Furthermore, many times, the order of the items are not known since they are not straightforwardly comparable. Also measurement errors might lead to responses that do not strictly fit the ordering. As a result we can no longer conclude deterministically that if a subject answers a question affirmative, whether she will be able to give affirmative answers to other questions of lower order in the same questionnaire. We use Rasch model to overcome this drawback by taking into account measurement error.

B. Rasch Model

Rasch model [4] provides a probabilistic framework for Guttman scales. In Rasch model, the probability of a specified binary response (e.g. a subject agreeing or disagreeing to an item) is modeled as a function of subject's and item's parameters. 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

¹<http://r-forge.r-project.org/projects/erm/>

this paper Rasch models are used to assess the organizations degree of being radical or counter-radical based on the religious-social keywords (items) appearing in their rhetoric.

Rasch model also 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 simple Rasch model the probability that subject v responds 1 for item i is given by

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

Rasch model assumes that the data under analysis have the following properties

- 1) *Unidimensionality*: $P(x_{vi} = 1 | \theta_v, \beta_i, \alpha) = P(x_{vi} = 1 | \theta_v, \beta_i)$, i.e., the response probability does not depend on other variable
- 2) *Sufficiency*: sum of responses contains all information on ability of a subject, regardless which item it has responded
- 3) *Conditional independence*: for a fixed subject there is no correlation between any two items
- 4) *Monotonicity*: response probability increases with higher values of θ , i.e., subject’s ability

Items with $s_i = \sum_v x_{vi}$ value of 0 or n , and subjects with $r_v = \sum_i x_{vi}$ value of 0 or k are removed prior to estimation, where n is the total number of subjects and k is the total number of items. Running Rasch model on the data gives us an Item parameter estimate or a score for each item. Generally the estimation of β_i or score for a item i is calculated through Conditional Maximum Likelihood (CML) estimation [11]. The conditional likelihood function for measuring item parameter estimate is defined as

$$Lc = \prod_v P(x_{vi} | r_v) = \frac{\exp(-\beta_i s_i)}{\prod_r \sum_{x|r} \exp(-\beta_i x_{vi})}$$

where r represents the sum over all combinations of r items. Similarly maximum likelihood is used to calculate subject parameter estimation θ_v or score for each subject.

In order to evaluate the quality of these measurements we run Anderson Likelihood Ratio test (LR-test) [7] on the set of data. The test gives us a goodness of fit of the data in Rasch model, i.e., it tells us whether the data follows the assumptions of Rasch model. A p -value, returned by the test, indicates the goodness of fit and a p -value² higher than 0.05 indicates no presence of lack of fit.

C. Implementing in Text Mining Domain

In this paper, we use Guttman scaling and Rasch model to find a ranking of some political organizations based on how extreme their views are on radicalism and counter-radicalism. In our project, model subjects are a group of religious organizations and items are a set of socio-religious *radical*

and *counter-radical* keywords. These keywords represent the radical and counter-radical beliefs and practices of the organizations. An organization responding “yes” to a feature means the organization exhibits that feature 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 *degree* of radicalism or counter-radicalism exhibited by an organization’s rhetoric. Details of keyword selection is presented in the next section.

III. METHOD

A. Problem Definition

The goal of this study is to build a semi-automated method to rank religious organizations from a certain geographical region on a scale of radicalism vs. counter-radicalism using their web sites. The efficacy of the generated model is evaluated by comparing it against baseline methods and expert level performance.

B. System Architecture

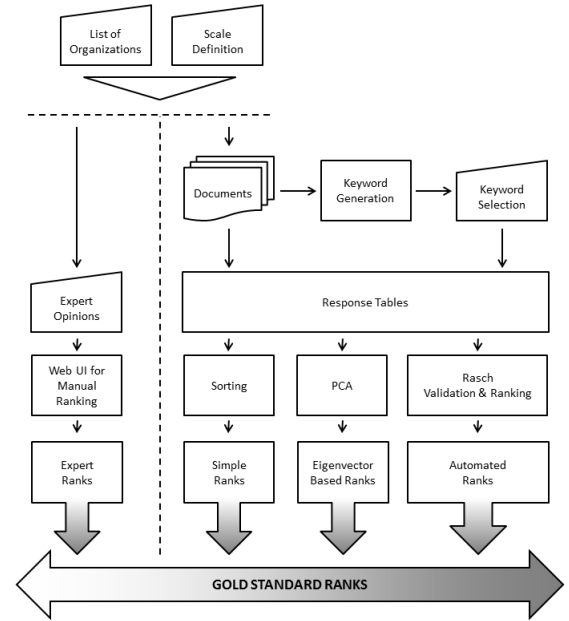


Fig. 1. A model of the system architecture.

The system architecture is shown in Figure 1. Here the flow of the processes and data can be seen as interactions between experts and automated modules. The system works as follows:

- Initially, social scientists use their technical and area expertise to identify a set of *organizations*, and hypothesize any number of unipolar or bipolar *scales* that could explain the variance among the beliefs and practices of the organizations. In this paper, we primarily focus on the construction and validation of the bipolar radical vs. counter-radical (R/CR) scale, however the techniques can

²<http://en.wikipedia.org/wiki/P-value>

be readily applied to the construction and validation of any other relevant scale.

- Next, we crawl and download the web sites of the organizations, and the system automatically *extracts the top-k candidate keywords* for consideration in the hypothesized scale. Social scientists screen the list of extracted keywords, and *select* the relevant ones for inclusion in further analysis.
- The system builds *response tables*; a pair of tables for a bipolar scale (such as R/CR), or a single table for a unipolar scale, by thresholding the occurrence frequencies of the selected keywords in the organizations' web corpus. See Fig. 3 and Fig. 4 for the response tables for the R/CR scale. The response tables are fed as input to the *Rasch Model building* algorithm. The algorithm produces a metric to *validate* the fitness of the model, and *rankings* of the organizations and keywords.
- In parallel to this, two types of other information are collected for *evaluation* purposes. First, *expert rankings* of the organizations, using a graphical drag-and-drop expert opinion elicitation tool shown in Fig. 2). Expert rankings are merged into a consensus *gold standard* of rankings. Next, two other computational baseline methods; one based on simple sorting, and another based on principal component analysis [12], are used to generate alternative *computational rankings* shown in Fig. 5.

C. Keyword Selection

In order to identify candidate keywords, one option was to translate the documents into English and apply readily available keyword extraction methods [13]. However it was preferable to preserve the original expression of the phrases in the original language. Therefore, we utilized a non-linguistic technique that relies only on statistical occurrence and frequency information.

Within each document, the words were separated by whitespace or punctuation marks. We considered each keyword to be an n-gram of one to three words. We treated each organization as one document and calculated the term frequency - inverse document frequency (TF-IDF) [14] values for every single n-gram mentioned by these organizations. The n-grams with highest TF-IDF value gave us the topics that each of the organization discusses most. Then the top 100 n-grams from each organization were made into a list of candidate keywords. Finally, belief and practice keywords that belong to one of the following categories {social, politics, economics, religion} were manually identified by the experts as relevant for inclusion. This process assessed a total of 790 candidate keywords; of which 29 and 26 were selected by experts for inclusion in the radical and counter-radical scales respectively.

D. Feature Extraction

After identifying the keywords for the analysis, we needed to search the web site corpus of the organizations for the matching items. This yielded a term-document matrix.

This task was performed in a simple three step procedure; initially the occurrence frequencies of particular keywords were counted within each organization's corpus, then a threshold matrix was calculated from the 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 (1), where k is the keyword, o is the organization, and D_o is the document set pertaining to that particular organization.

$$f_{o,k} = \frac{|\{d \mid k \in d, d \in D_o\}|}{|D_o|} \quad (1)$$

A threshold value for each keyword is calculated by taking the median of the values in the related column. Median was preferred over mean as a threshold, since the distribution of the values did not fit Gaussian distribution, yet median empirically proved to be a better measure.

Finally, each element was converted into a binary value by comparing it to the column's threshold. English translations of the keywords is presented for clarity in Fig. 3 and Fig. 4, additionally names of the corresponding organizations are anonymized consistently in all figures and tables.

E. 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 package³, to fit a Rasch model to the dataset, and obtain the organizations' scores on the latent scale, which are the the subject parameter estimates (θ_v) discussed in previous section. The eRm package⁴ fits Rasch models and provide subjects or organizations parameter estimates based on maximum likelihood estimation.

The automated scale of the organizations is formed by ranking the organizations according to their estimates on the latent scale. Not only we can provide the organization estimates but we can also assess whether the model fits the data by looking at several goodness of fit indices, such as the Andersen's *LR*-test.

IV. EXPERIMENTAL EVALUATION

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 2005 to 2011. For each web site, a specific REGEX filter was used to strip off the headers, footers, advertising sections and to extract the plain text from the HTML code.

³<http://cran.r-project.org/>

⁴<http://r-forge.r-project.org/projects/erm/>

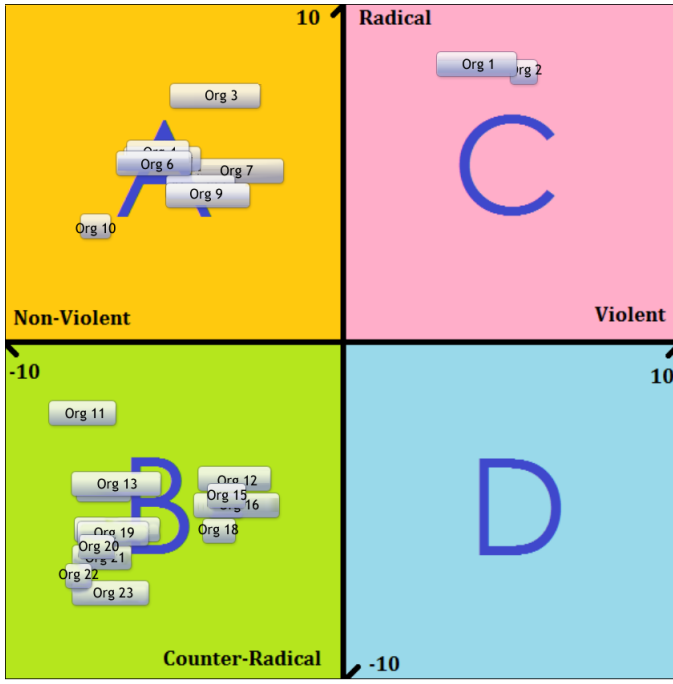


Fig. 2. The visual interface of the expert opinion collector for manually placing the organizations on the two dimensional scale

B. 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. In order to build a gold standard of orderings of the organizations, we built a graphical drag-and-drop user interface tool to collect the opinions of each of the area experts. A screenshot of the tool is shown in Fig. 2.

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. Also, 90% of the items have less than 22.6% difference in their rankings. The individual scores for each organization were combined and averaged to obtain the consensus *gold standard rankings* along the hypothesized R/CR scale.

C. Computationally Generated Scale

The ranking discovered by the Rasch model fitting the corpus has been evaluated against the gold standard rankings of the organizations provided by the experts. The difference between two separate rankings have been calculated by using the following misplacement error measure in Equation 2.

$$error(G, R) = \frac{\sum_{o \in O} \frac{|G(o) - R(o)|}{|O|}}{|O|} \quad (2)$$

Here, O is the set of organizations, G and R are one to one mapping functions of rankings from set O to range $[1, |O|]$.

	RADICAL	Quran	Islamic teachings	Shari'a	infidel	military strength	Shari'a	Caliph	prophet muhammad	prophet muhammad	Tabligh	Sharia	corruption	Shirk	paganism	street preaching	obligation for Shari'a	Wahiddeen council	the Zionist Jews	Islamic thoughts	journalism	Shah council	Wahiddeen	suicide bomb	communists	the delators	Sharia enforcement	violence	Islamic state
Org 9		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Org 7		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Org 1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Org 3		1	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Org 4		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Org 8		0	0	1	0	1	1	1	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	
Org 2		1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	
Org 5		0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
Org 6		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Org 10		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Fig. 3. Radical subset of organizations and keywords, sorted according to aggregate row values.

	politics	equality	hindu	pluralism	tolerance	gender justice	human rights	indonesian islam	religious freedom	intercultural dialogue	constitutional rights	democracy	election	christian	non-muslims	judicial society groups	liberal	secularism	activists	peace building	conflict prevention	freedom of expression	homosexual	multicultural education
Org 12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Org 21	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Org 15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Org 11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Org 17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Org 14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Org 16	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Org 19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Org 18	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Org 20	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Org 13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Org 22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Org 23	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 4. Counter-Radical subset of organizations and keywords, sorted according to aggregate row values.

For two exactly matching rankings, the $error(G, R)$ will be zero, whereas for two inversely sorted rankings it is expected to be 0.5 (when the size of O is even). Also a random ranking is expected to have a error of 0.375.

D. Expert-to-Gold Standard Error

We calculated the error between each expert's ranking and their consensus gold standard of rankings. The first expert's error measure is 0.06, and the second and third expert's errors are 0.12 and 0.14 correspondingly as shown in the last row of the table in Fig. 5. The average error of our experts against their gold standard ranking is 0.11.

E. Baseline - Sorting with Aggregate Score

The first baseline we used was constructed by sorting the organizations according to the number of different keywords observed in their corpus. While this provided a pattern similar to a Guttman Scale, and orderings of the organizations matched to a certain degree with the gold standard as shown in Fig. 5, the $error$ for this baseline was 0.19, which is higher than the average expert's performance.

F. Baseline - Principal Component Analysis

A stronger baseline was built by employing principal component analysis [12], and sorting the organizations according to their projections in the first principal component of the term-document matrix. Since experts selected the R/CR scale relevant keywords only, it was expected that the first principal component would reflect the corresponding scale. PCA proved to be performing better than the aggregate score sorting, with an $error$ measure of 0.18. However, this error rate is still higher than the error rate of each expert.

G. Performance of the Rasch Model Ranking System

The p -values from the Anderson LR goodness of fit test from model (1) and model (2) (mentioned in section III-E) are 0.85 and 0.669 respectively, suggesting no evidence of lack of fit. The Rasch models allow us to get a natural order of the organizations, according to their “abilities”, i.e.: radicalism and counter-radicalism in this case. This system had an *error* measure of 0.10, which actually provided a higher ranking performance than the average performance of our experts’ – performing better than the majority of our area experts.

H. Evaluations

Our experiments showed that the hypothesized compatibility of the R/CR scale for the Indonesian corpus is valid. Not only the Rasch model was statistically fitting the response matrix, but also the generated ranking performance was better than the average expert performance. Among our computational baseline methods, the Rasch Model was the only method producing expert-level performance as shown in Fig. 5. This preliminary analysis with the R/CR scale shows that when experts assist the system with keyword selection, the web corpus of organizations provides rich enough information and patterns to enable a computational method to rank them accurately.

V. CONCLUSIONS AND FUTURE WORK

In our experiments, not only did the data show fitness with the Rasch Model for the R/CR scale, but also the Rasch rankings of the organizations are better than the output of the other baseline computational methods, and they are at expert level performance when compared with the consensus gold standard rankings.

Rasch Model also provided us with another output, namely the ranking of selected keywords (items) on the R/CR scale. Although preliminary observations indicates that this can be a valuable asset by itself, we plan to further investigate the quality and utility of this ranking as future work.

While the model has been demonstrated to fit on the R/CR scale, two major expansion points can be investigated in the future work, namely the violent/non-violent scale, and enhancement of feature selection. Although our experts have identified a second dimension, evaluating its correlation to R/CR axis, or existence of other significant ones could be beneficial. Additionally, the features can be enhanced by experimenting with the significance of the radical keywords in the counter-radical organization corpuses, and vice-versa.

Other interesting work includes making our expert opinion elicitation tool available online to a wider and more geographically distributed audience to crowdsource [15] the needed expertise for making lists of local organizations, identifying their web sources, and overcome the complex task of construction and validation of significant and fitting scales. Another interesting dimension is to look at synthesis and analysis of scales that do have a strict hierarchy of keywords, but adhere to more flexible partial order models [16].

	Computational Rankings				Gold	Expert Rankings		
	Random	Sort	PCA	Rasch		Expert 1	Expert 2	Expert 3
Organization Rankings	21	9	1	6	1	2	7	1
	15	7	7	5	2	3	2	5
	22	1	9	2	3	1	4	8
	19	3	4	8	4	4	6	3
	4	4	12	4	5	5	3	4
	10	8	3	3	6	6	1	9
	6	2	8	1	7	8	5	2
	2	5	2	7	8	9	9	10
	17	6	17	9	9	7	8	7
	5	10	5	12	10	10	10	6
	11	23	21	21	11	14	13	12
	20	22	15	15	12	12	11	21
	13	13	11	11	13	11	19	16
	7	20	14	17	14	17	18	15
	3	18	16	14	15	15	15	19
	23	19	6	16	16	18	20	20
	1	16	20	19	17	13	22	14
	14	14	19	18	18	16	17	11
	12	17	18	20	19	22	14	18
	9	11	22	13	20	19	16	17
	16	15	13	22	21	20	23	23
	18	21	10	10	22	21	12	22
	8	12	23	23	23	23	21	13
Error	0.36	0.19	0.18	0.10		0.06	0.12	0.14

Fig. 5. Computational and expert rankings

REFERENCES

- [1] R. Crellin, “Analysing terrorism and counter-terrorism: A communication model,” *Terrorism and Political Violence*, vol. 14, pp. 77–122, 2002.
- [2] H. Davulcu, S. T. Ahmed, S. Gokalp, M. H. Temkit, T. Taylor, M. Woodward, and A. Amin, “Analyzing sentiment markers describing radical and counter-radical elements in online news,” in *Proceedings of the 2010 IEEE Second International Conference on Social Computing*, ser. SOCIALCOM ’10. IEEE Computer Society, 2010, pp. 335–340.
- [3] L. Guttman, “The basis for scalogram analysis,” *Measurement and prediction*, vol. 4, pp. 60–90, 1950.
- [4] D. Andrich, *Rasch models for measurement*. Sage, 1988, no. 68.
- [5] R. D. McPhee and S. Corman, “An activity-based theory of communication networks in organizations, applied to the case of a local church,” *Communication Monographs*, vol. 62, pp. 1–20, 1995.
- [6] L. Le Cam, “Maximum likelihood an introduction,” *ISI Review*, vol. 58, no. 2, pp. 153–171, 1990.
- [7] D. Hessen, “Likelihood ratio tests for special rasch models,” *Journal of Educational and Behavioral Statistics*, vol. 35, no. 6, p. 611, 2010.
- [8] R. Likert, “A technique for the measurement of attitudes,” *Archives of Psychology*, vol. 140, pp. 1–55, 1932.
- [9] L. L. Thurstone, “Attitudes can be measured,” *American Journal of Sociology*, vol. 33, pp. 529–554, 1928.
- [10] J. McIver and E. Carmines, *Unidimensional Scaling*. Sage Publications, Inc, 1981, vol. 24.
- [11] Y. Pawitan, *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press, USA, 2001.
- [12] I. Jolliffe, *Principal Component Analysis*. Springer Series in Statistics, 2002.
- [13] W. Michael and J. Kogan, *Text Mining: Applications and Theory*. Wiley, 2010.
- [14] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” in *Information Processing and Management*, 1988, pp. 513–523.
- [15] R. Snow, B. O’Connor, D. Jurafsky, and A. Y. Ng, “Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks,” in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, ser. EMNLP ’08. Stroudsburg, PA, USA: Association for Computational Linguistics, 2008, pp. 254–263. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1613715.1613751>
- [16] A. W. James and L. M. John, “Algebraic representations of beliefs and attitudes: Partial order models for item responses,” *Sociological Methodology*, vol. 29, pp. 113–146, 2002.