SANDIA REPORT

SAND2009-4953
Unlimited Release
Printed May 2009

Supersedes SAND2009-3135
dated May 2009

Using Adversary Text to Detect Adversary Phase Changes

Adele B. Doser, Ann E. Speed, Christina E. Warrender

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND 2009-4953
Unlimited Release
Printed May 2009

Using Adversary Text to Detect Adversary Phase Changes

Adele B. Doser, Ann E. Speed, Christina E. Warrender

Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185

Abstract

The purpose of this work was to help develop a research roadmap and small proof of concept for addressing key problems and gaps from the perspective of using text analysis methods as a primary tool for detecting when a group is undergoing a phase change. Self-organizing map (SOM) techniques were used to analyze text data obtained from the world-wide web. Statistical studies indicate that it may be possible to predict phase changes, as well as detect whether or not an example of writing can be attributed to a group of interest.

Table of Contents

Abstract 3
Table of Contents 5
List of Figures 5
List of Tables 5
Introduction 7
What is phase change? 7
Other work in this area 7
Selection of study group 8
Data Collection 9
Initial analysis via SOM 10
Data sets 10
Data processing 11
Results using test data unseen by the SOM 13
What can we generalize from this? 14
Statistical Analysis of Data Set and SOM Technique 14
Bootstrapping 15
Jackknifing 19
Summary and Conclusions 21
References: 25
DISTRIBUTION 26

List of Figures

Figure 1: Left: U-Matrix of SOM, right: resulting groups after clustering 12

List of Tables

Table 1: Boostrapping results. Seed group was part of training set (constrained) 16
Table 2: Bootstrapping results. Seed group may or may not have been part of training set (unconstrained) 18
Table 3: Jackknife results 20

Introduction

What is phase change?

In chemistry, “phase change” denotes the transition of a mixture from one phase to another without a change in chemical composition. For example, ice melting into water undergoes a phase change. In mathematics, “phase” is used to denote the position of a periodic signal. “Phase change” is often used in mathematics to mean that a periodic signal has changed from monotonically increasing to monotonically decreasing, or vice versa. For example, when the angle of a sine function increases from 0 to $\pi/2$, its value increases from 0 to 1. As the angle increases from $\pi/2$, the value of sine begins to decrease. Thus, at the point of $\pi/2$, a sine function is said to undergo a “phase change”.

In our study, it is the latter mathematical definition that is more applicable. When we refer to “phase change” we mean that a group has undergone a fundamental change in philosophy or attitude. Philosophical changes are items such as ideology, intent, and sentiment. These changes will often manifest themselves as changes in action, such as a formerly peaceful group suddenly adopting violence as part of their dogma. The changes include agreements to ceasefires, violations of ceasefires, or desperate actions of a losing fight such as a sudden adoption of suicide bombing techniques.

We explore the notion of phase change. Specifically, this report attempts to answer the question: is there some signal that can be monitored in the writings of a group that will signal an imminent change of phase? To answer this question, we gathered information on world organizations advocating regime change, both violent and peaceful. We selected a group for which we had evidence of phase changes (e.g., transitions from non-violence to violence or upswings in violence or changes in tactics), and had sufficient data for analysis. The writings of the group were collected and analyzed through use of self-organizing maps (SOMs). Various signatures, including signatures of phase change, were found in the data.

Other work in this area

In previous work by Allison Smith [8], she studied the psychological characteristics and rhetoric of organizations operating in the same time and place. These analyses were based on match pairs in which each pair consisted of one future violent group, one non-violent group. She studied documents issued by seemingly similar group pairs to determine if the future terrorist acts of the violent group could have been predicted. She rated groups on variables such as “power motive” and “practicality value”. These values were somewhat subjective and manually coded by a group of scorers. She asserted that content analysis could be used as a predictor of future behavior.

Stephan Green [2] took a similar approach to Smith, but concentrated on detecting sentiment in text. While Green’s approach was largely automated and achieved results much better than baseline, it was based upon grammatical relationships which were defined a priori.

Steven Shellman [5-7] has successfully modeled phase change of terrorist organizations based on measures such as cooperative and conflictive actions. Unlike the other studies mentioned, he concentrated on actions, rather than intent. His work was based upon news articles about groups, rather than the writings of the groups themselves.

Our work differs from these approaches in several ways. First, we assume nothing a priori about the group's use of language. We do not manually look for sentiment in text, for instance. Other than a manual inspection of a word and phrase list to eliminate redundancies and selection of two or three writings of interest, our method is completely automated. These activities, too, could be automated in the future. Finally, our training method is un-supervised. We let the algorithm decide what is important in the text to determine how a document should be categorized.

Selection of study group

Many groups were identified as good candidates for this study. Ideally we wanted to compare the analysis of a group that underwent a phase change with a similar group operating in the same general area/time with similar goals that did not exhibit a phase change. Several potential group pairs were obtained from Steve Shellman and literature such as Allison Smith's dissertation. Many of the control groups listed in Allison Smith's dissertation were later determined by the US government to be funneling monies to terrorist organizations and thus could not be considered as controls. Many others, such as the Palestinian authority, quit writing when they achieved many of their goals and their writings are not archived. Many of the Indonesian groups do not write in English. When one passes the groups through these filters, not many good candidates remain. We did identify a group pair that looked to be an excellent candidate based on language, activities, and quantity of writing. This pair consisted of Operation Save America (non-violent US abortion activists), and Operation Rescue (sometimes violent US abortion activists). But due to Sandia's close relationship with the government, we were cautioned away from this option since it involved the direct study of US citizens.

It was decided to pursue the Liberation Tigers of Tamil Eelam (LTTE) of Sri Lanka, commonly referred to as the Tamil Tigers, as a study group. For more than 25 years, the LTTE has been engaged in a civil war with the government of Sri Lanka. The island of Sri Lanka is approximately 80% Buddhist, and 12% Tamil Hindu. Many Hindus feel that they are persecuted by the Buddhist majority. The Tamils were peaceful for decades. Then, in July 1983 there were several days of rioting where it was reported that Buddhist Sinhalas attacked and killed as many as 3,000 of the Tamil minority. The government and the police were accused by many news organizations of turning a blind eye. This event allowed the LTTE to gain power and begin their quest for Tamil independence. They instigated many types of activities that were later adopted by Al Qaeda and others. For instance, LTTE was reported to be the first terrorist group to use female suicide bombers.

In early 2002, an uneasy ceasefire was signed between the LTTE and the Government of Sri Lanka (GoSL). There were many phase changes exhibited by the Tigers before the ceasefire fully fell apart four years later. In addition, there was also a brief ceasefire in late 2000. Thus, the period from 2000-2006 would be good to examine. The LTTE maintained a web site in English directly attributed to them, which contained not only news stories, but dogma and sentiment as well. News on the site was highly influenced by the underlying tenets of the LTTE. Articles on the LTTE web site were archived for the period of time in question. There were large Tamil diaspora communities in Canada, the UK, and Singapore. Many of them were sympathetic to the LTTE and also wrote in English. Steve Shellman suggested there was a sister group, the TULF, that would be a good control. The LTTE seemed to have all the characteristics we were looking for. Thus, we selected them as our study group. Unfortunately, we later found that the TULF either disbanded or was absorbed by the LTTE and their writings were not archived.

The remainder of this report is organized as follows. The next section discusses the collection of web documents containing writings by the LTTE. The following section discusses the analysis of these writings using a self-organizing map (SOM) technique. Next, the SOM analysis is repeated using various statistical techniques to determine the mathematical merit of the data set. Finally, a summary and conclusions are provided.

Data Collection

There were two issues involved in the data collection for this project. The first was the accumulation of potential writings that may have been relevant. The second involved sifting through these documents to determine whether or not they were useful.

Data collection for the project was conducted on a home computer using a free web-based spider. Seeded with URL's of what we knew to be LTTE writings, the spider collected thousands of documents. Many of these were about the Tigers. However, if an article about the Tigers compared Tamil persecution to the Jewish holocaust, the spider would then start collecting articles about the Jewish holocaust. Thus many collected articles were not about the Tigers. For operational security reasons, we chose not to use the Stanley spider. If we would have, we might have narrowed the collection down to a small number which were potentially useful. This operation largely had to be done manually.

There are a number of difficulties incurred in the attribution of writings to a particular group or even a sympathizer. As an example unrelated to the LTTE, in the initial search for appropriate study groups a blogger was uncovered who posted many articles in 2008. He claimed to be from Jemaah Islamya and asserted the group was endorsing Barak Obama for president. After a couple months, the blogger admitted that he was not from Jemaah Islamya, but was an American trying to sway public opinion away from Obama. There are many examples of false writings such as this on the web. Had we used the

Stanley spider, it is unlikely it would have flagged these early blogs as inappropriate. There would have been some human analysis required even in the best case scenario.

Much of the data collected from the free spider were news articles posted in the various Tamil diaspora sites around the world. After a manual search, we were able to obtain a small nucleus of documents that we were either certain the Tigers had written, or they were clearly written by Tiger sympathizers. Roughly, there were 100 files we knew the Tigers had written. These were used for training and testing purposes. There were another 40 documents used solely for testing which included not only articles from Tamil Tiger sympathizers, but control documents such as a posting from the Sri Lanka government and an article about Tamil cuisine. Ultimately, we would have liked to have collected more relevant documents. However, we felt we had enough to perform a preliminary analysis and produce some useful results.

Another complication made data collection more difficult. In early 2009, the LTTE suffered major setbacks in their war with Sri Lanka. In January, the LTTE capital of Kilinochchi fell to the Sri Lankan government army. A few weeks later, the main LTTE website, where much of the data was collected for this study, was taken offline. A partial archive had been made by an independent group and was accessed, but it was poorly indexed.

Initial analysis via SOM

A self-organizing map (SOM) analysis was selected to be the analysis tool. We ran the SOM on a nucleus of available data to see if it would produce any useful results. That study is described in detail below.

Data sets

For the training data, 63 web pages attributed to the Tamil Tigers were used. The web pages contained a variety of sentiments, including party propaganda, news reports, speeches from the LTTE leader, discussions of the peace process, childhood welfare, denials, accusations, and ultimatums. In the training set, the contents of two web pages could be strongly connected to a subsequent attack by the Tamil Tigers. For example, in one posting, a mention that the LTTE was going to intensify its struggle was followed by three attacks over the following weeks during a period of time when a ceasefire was supposedly in place.

The testing data set consisted of 33 web pages taken from various sites with strong connections to the LTTE, or from news sites containing interviews with LTTE leaders. Some pages demonstrating support for the LTTE but lacking a firm connection to the Tigers were used, such as blog entries or editorials. For control, documents which were obviously unrelated were used, such as a web page that discussed Tamil cuisine.

In a real scenario, one would like to insure that documents truly attributable to the group of interest were used for training purposes. This philosophy would change once the tool is trained. In a typical day to day situation, one may not have time to verify the origins of an incoming stream of data. Thus we included testing data from various sources.

Data processing

A self-organizing map approach was selected for this study due to its long reputation as a useful pattern recognition tool. The textual data needed to be processed to a format that would be useful to MATLAB, the program used to run the SOM algorithms. The web pages were processed through the Sandia Analyst Aide, resulting in a matrix containing key words or phrases for each web page in the training data. The result was approximately 5000 key words for the training data set. Redundancies were eliminated and the list was whittled down to approximately 1400 words or root words. For example, "violent" and "violence" are words that express the same sentiment, and can both be expressed by the root "violen".

A MATLAB script was written to compare the words in the web pages to the list of key words, resulting in a numerical matrix which expressed the number of times a word, root word, or phrase appeared in each web page. Finally, the matrix was normalized so that all the web pages would carry the same weight.

Legacy code was utilized in the analysis of the resulting matrix data. A modified self organizing map (SOM) developed for another project was employed. By itself, a SOM will not classify data, but it allows for the visualization of multidimensional data in a lower dimensional space, while still preserving the distance relationships among the data points. For example, if the parameter space has 14 dimensions, after processing with a SOM, data points which were far away from each other in 14 dimensional space will still be far apart in 2 dimensional space. Because of the dimensionality reduction involved, the SOM mapping is not unique. That is, one data point in the finished SOM will likely represent more than one data point in the original data set. Based on the eigenvalues of our training data set, the SOM algorithm came up with a map that had dimensions of 5x8 cells. That is, each web page could be mapped to one of 40 locations in the SOM.

The resulting SOM representation can be classified to put similar data into groups. The legacy code used allows the user to "seed" the classifier with the location of data points that are clearly different. For instance, a web page discussing child protection issues, and a web page expressing accolades for a suicide bomber should clearly belong in two different groups. A modified K-means clustering algorithm uses the seed locations as starting points for classification, and based on centroid distances, the result of the clustering technique contains at least as many groups as seed files, usually more. The user then has the opportunity to examine the results and merge groups if desired.

In this case, three seed file locations were provided, a web page containing an ultimatum, a web page discussing children's issues, and a web page where the LTTE is denying they committed an act the government has attributed to them. Based on the locations of the data centroids, the classification algorithm produced 5 distinct groups. The U-matrix (Unified distance matrix) is shown on the left. The U-matrix visualizes the distances between adjacent units in the SOM. Red and brown represent areas where the distances between the nodes are large, and blue areas are where the distance is small. The U-matrix gives one an idea how many natural clusters are present in the data.

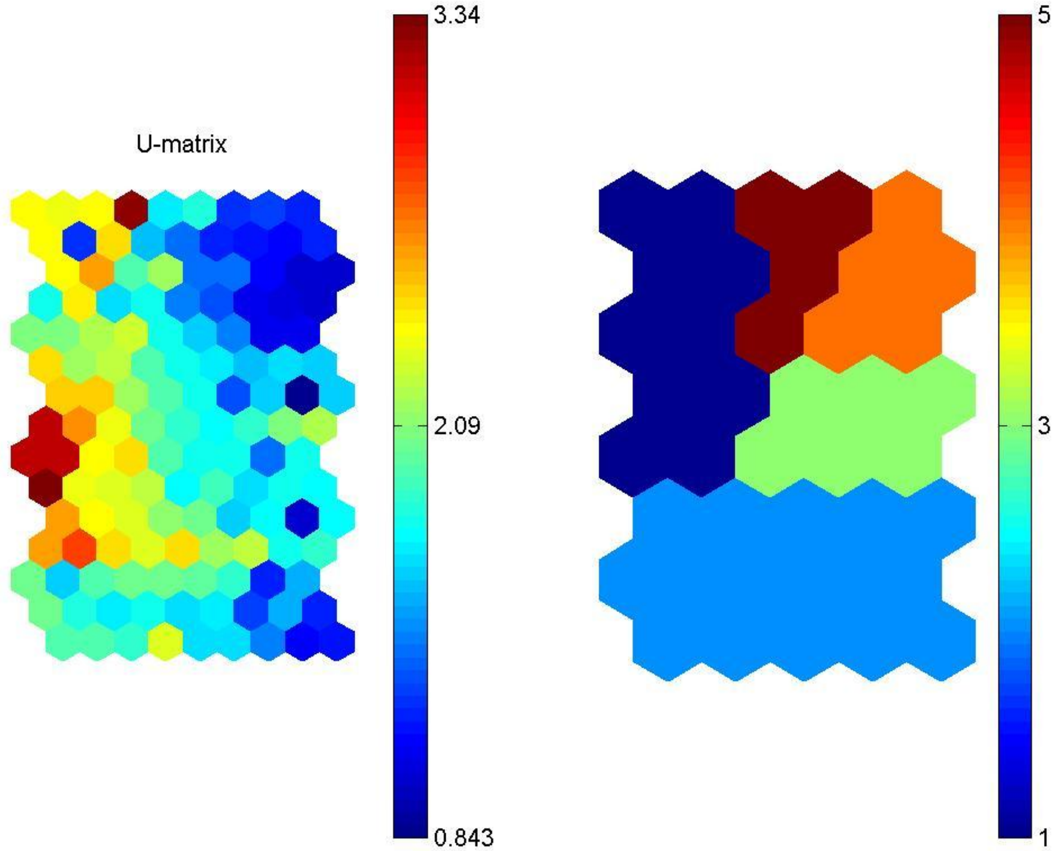

Figure 1: Left: U-Matrix of SOM, right: resulting groups after clustering

Generally speaking, the data that map to the blue areas are not as linguistically diverse as the data that map to the red and brown areas. The right side of the figure represents the result of the clustering. Given three seed points, five distinct groupings resulted. The colors and numbers attributed to each class have no inherent meaning. When compared to the training data, the following pattern emerges.

1) The dark blue cluster was essentially a potpourri of files that didn't fit well into other groups. An editorial about the aims of the new Sri Lankan president and a discussion of ethnic cleansing fell into this group. Note that comparison with the U-matrix reveals that this class contained the most diverse data.

2) The light blue cluster contained files with inflammatory language, ultimatums, and many speeches from the LTTE leader. *Both web pages connected to a subsequent LTTE attack fell into this data set*.
3) The green cluster contained pages that could be largely classified as news articles. That is, reporting of events rather than commentary.
4) The orange cluster contained files that discussed children's issues, humanitarian issues, and the peace process. In general, things not related to attacks.
5) The brown cluster contained files where the LTTE was either reacting to actions taken by the Sri Lankan government, or the LTTE was denying attacks that others had said they had executed.

It is also interesting to note that files related to the peace process (group 4), and files related to ultimatums, etc.(group 2), were the least linguistically diverse.

There were a few articles in this set that on the surface appeared to be inflammatory, and the SOM did not place them in the inflammatory group (2). However, when they were correlated with the timeline of events, one could see that the Tigers committed no attacks that year. So perhaps the events were not misclassified. There were also a number of web pages classified in the inflammatory group that did not precede an actual event.

Perhaps the most interesting finding from the training data set had little to do with phase change on the part of the LTTE. I found a couple of examples where an LTTE web page placed in the peaceful grouping (4) preceded an attack by the other side. So perhaps peaceful dialog on the part of the LTTE is perceived by the Sri Lanka government as a sign of weakness.

Results using test data unseen by the SOM

The testing data came mostly from sources other than the official LTTE web sites and did not adhere as strictly to the nice groupings above; however, four documents did come from the official LTTE web sites. The following results were obtained.

a) There were three files that discussed children's issues or the 2004 tsunami. They all mapped to group (4).
b) The article on Tamil cuisine mapped to group (3), the news article group. While we were surprised this didn't map to the potpourri group, at least the SOM didn't think Tamil cuisine was an imminent threat to peace.
c) There were four editorials from Tamil supporters in Canada. They all mapped to group (4), the children's issues and peace group.
d) There were 8 web pages of interviews with the leader of the LTTE. These all mapped to groups (2) and (3). However, the outcomes here could be biased by the text in the interviewer's questions. It might be useful to strip off the interviewer's questions and reprocess this text.
e) An article about female LTTE cadres from the same site where the training data was obtained mapped to group (3).
f) There were 3 blogs from Tamil supporters. Two mapped to group (3), one to group (4).

g) One article was taken from an official web site of the Sri Lankan government. This article denied the government's involvement in acts the LTTE attributed to it. The web page mapped to group (5), the denial group.

A few interesting notes: there were no files that mapped to the potpourri group. In addition, no files that could not be attributed to the LTTE leader mapped to group 2 (the inflammatory speech group).

What can we generalize from this?

The SOM analysis technique seemed to do a good job at spotting polarized writings (violent/peaceful) even though examination of the individual pages revealed that actual word use was not a good indicator. That is, when a file that mapped to group (2) was compared with one that mapped to group (4), there was no one word or small subset of words that emerged as a discriminator. While that is useful, we are more concerned with the predictive nature of the SOM. The underlying problem is relating the time line to the data set.

A subjective assessment of the data collected was conducted by the project authors, who are experts in various areas of information retrieval and pattern recognition but are not trained intel-analysts. Many terrorist acts were pinned on the LTTE by the Sri Lankan government that did not seem to be supported by evidence and the LTTE strictly denied having performed them. We tried to concentrate on actions such as political assassinations and suicide bombings that were very obviously carried out by the Tigers. Two web pages were determined manually due to their sentiment and time placement to strongly correlate with subsequent LTTE attacks after a peaceful period. These articles both mapped to group (2). No article outside group (2) could be correlated with a subsequent terrorist event.

In addition, the SOM seems to give us an idea if a document expressing violent sentiment was actually written by the Tigers. Although many of the sympathizer documents in the testing data set exhibited hateful writing, none of them were actually mapped by the SOM to the dangerous group (2). These results would seem to suggest that if a web page actually mapped to group (2), we could be confident it was authored by the Tigers, and be relatively sure that they mean business.

Statistical Analysis of Data Set and SOM Technique

Due to the limited size of the data set, analysis via bootstrapping and jackknifing techniques was conducted to support the credibility and impact of the research.

Bootstrapping

In bootstrapping analysis, some number N of samples from the original data set is drawn at random, with replacement, and the functional estimate (in this case a SOM) is performed. The process is repeated M times. The results are analyzed to determine useful statistics such as confidence intervals in the case of fuzzy sets or percentiles in the case of discrete sets. It is the nature of the bootstrap that every run will produce slightly different results. Since the samples are drawn with replacement, there is a small risk of running into invertibility problems if too many copies of the same sample are used. Fortunately, that problem did not occur in our experiments.

We selected a value of 63 for N, which was the number of files used in the original SOM analysis described above. We had a total of 100 files (web pages) that we could directly attribute to the LTTE. All 100 files were used for the potential training set. That is, 63 files, with replacement, were selected from a total data set of 100.

M was selected to have a value of 200. Thus, 200 different draws of 63 data files each were run through the SOM.

The SOM training was just one facet of the algorithm. It was also necessary to cluster the resulting U-matrix into groups. As described in the previous section, in order to perform the clustering, the user would provide a number of "seed" files from the data which should clearly belong to separate classes. The requirement presents a dilemma. Technically, all N samples from the original data set should be selected at random. If we pre-select a number of files (three in this case), we are violating that rule. However, the seed files should ideally be from the training set, not the testing set. Ultimately, we made the decision to run the bootstrapping analysis twice. The first 200 runs were with the three seed files being forced to be part of the training set (constrained). In the remaining 200 runs the seed files, due to the nature of the random selection, would sometimes be in the training set and would sometimes be in the testing set (unconstrained). Since the SOM may or may not have seen the seed data during the training phase, we would expect the results of the second set of bootstrapping runs to be slightly worse.

Different data have different eigen values. Since the size of the SOM is based on the eigen values of the data, potentially every boostrapping run may produce a SOM of a different size. In addition, the seed files, while constrained to be in different groups, will be in different locations from run to run. There may also be a different number of groups resulting from each run, depending on the data files selected. Many draws of the same data file for the training set will produce a less diverse SOM. Given all these realities, how do we measure the results?

There were a number of files that, based on manual interpretation and the early SOM analysis, clearly belonged in a certain group. For example, the inflammatory writings that preceded an attack definitely belong only in the group with other inflammatory writings. Writings that are not attributable to the Tigers should never belong in the inflammatory group. The seed set for clustering consisted of three files. The resulting number of groups may consist of three, but it may consist of five or six. Since the “extra” groups change from run to run, we cannot say much about them with certainty. However, if we restrict ourselves just to the three seed groups, and data that clearly belong to one of them, we have a way to measure the results. For example, were the files that clearly belong in the inflammatory group actually placed there in all the runs?

A subset of 12 files, both from the training and testing set, was selected that strongly related to one of the three seed groups. Performance was measured on how often out of the 200 runs the data sample was placed in the designated group. For demonstration purposes, a file whose group was not so clear was also included. It is designated below with an asterisk (*).

For consistency, the seed files are grouped as follows.

Group 1: News or denial (the seed file reported an attack on GoSL army truck by an unidentified group)
Group 2: Humanitarian or social issues (the seed file was an article about children’s issues)
Group 3: Inflammatory articles that were attributed to the LTTE (seed file was an ultimatum)

If a document was judged by the algorithm ***not*** to be in one of these groups a portion of the time, it is documented below as ***other***.

We would expect the constrained results (Table 1) to be slightly better than the unconstrained results (Table 2). In the constrained case, we are forcing the SOM to include the clustering seed files as part of the training set. In other words, we are telling the algorithm, “I want you to put these three files, which you have seen before, into different groups.” In the unconstrained case, we are telling the algorithm, “You may or may not have used these files to build your map, so they may be unknown to you. You may not have a location on your map that perfectly describes them. But I want you to put them in different groups.”

Table 1: Boostrapping results outlining portion of runs SOM actually placed file with presumed group. Total number of runs=200. Seed group was part of training set (constrained)

Data file description	Presumed Group via manual and old SOM	Actual group(s) found by boostrapping	Percent of runs in each group
Training data			
Ultimatum preceding attack	3	3	100
News about LTTE working to peace	1	1	100
Thimpu declaration (objectives of LTTE)	1	1/2	99.5/05
Article about peace	1	1/2/other	76/24/1
Article about killing of humanitarian workers	2	2/other/3/1	74/19.5/6/0.5
Speech from LTTE leader (2004), inflammatory language	1	3/2/1	93.5/3.5/3
Speech from LTTE leader (1992) more conciliatory	1 or 3*	1/3/2	52.5/36.5/10.5
Testing data			
Article about 2004 tsunami	1	1	100
LTTE sympathizer	1	1/2	95.5/4.5
Blog accusing Sinhala Buddhists of racism	1	1/2/3	85.5/11/3.5
Tamil cuisine article	1	1	100
Denial of suicide attack by sympathizer	1	1/2	79/21

*While the 1992 speech was more conciliatory than the 2004, from the language it could go either way. This file was included to demonstrate that a SOM can mimic the uncertainty in human decision making.

Table 2: Bootstrapping results outlining portion of runs SOM actually placed file with presumed group. Total number of runs=200 Seed group may or may not have been part of training set (unconstrained)

Data file description	Presumed Group via manual and old SOM	Actual group(s) found by boostrapping	Percent of runs in each group
Training data			
Ultimatum preceding attack	3	3	100
News about LTTE working to peace	1	1/other	99/1
Thimpu declaration (objectives of LTTE)	1	1/2/other	98/1/1
Article about peace	1	1/2	72.5/27.5
Article about killing of humanitarian workers	2	2/other/3/1	70.5/18/11/0.5
Speech from LTTE leader (2004), inflammatory language	1	3/1/2/other	91.5/6/2/0.5
Speech from LTTE leader (1992) more conciliatory	1 or 3*	1/3/2	52.5/39.5/7.5
Testing data			
Article about 2004 tsunami	1	1/other	99/1
LTTE sympathizer	1	1/2/3/other	95/2/2/1
Blog accusing Sinhala Buddhists of racism	1	1/2/3/other	86/9/4.5/0.5
Tamil cuisine article	1	1/other	99/1
Denial of suicide attack by sympathizer	1	1/2	73/27

(All training data was selected from a possible set of 100 files. Thus, in the "training data" results in the tables above, the file specified was in the training set, but may not have been selected by the algorithm for training purposes. For this reason, we would expect the algorithm not to obtain 100% correct classification for all files. In no case was the "testing data" viewed previously by the SOM.)

Overall, the SOM performance varied depending on what file it was given. It performed perfect in some situations, but worse in others. This implies that perhaps SOM performance is based more on quality of data rather than on quantity of data. It seemed to do well at spotting ultimatums. In fact, in the two examples we have of statements preceding an attack, the SOM put them in the exactly the same cell (point on the SOM map) 69% and 68% of the time in the constrained and unconstrained runs, respectively.

In the case of inflammatory writings (group 3) 20 files from the training set of 100 were placed in this group over 90% of the time and 27 files over 50% of the time in the constrained run. In the unconstrained run, the numbers in group 3 were 20 files for 90% of the time and 25 files 50% of the time. Our initial SOM training exercise in the previous section placed 23 files in this set. The bootstrapping exercise and the initial SOM exercise would seem to be in agreement. None of the test files were placed in group 3 with over 50% certainty. This finding was desirable, since group three was intended to consist of inflammatory writings attributed only to the LTTE.

As predicted, the bootstrapping analysis in which the seed files were not constrained to be part of the training set (the second table) performed more poorly. However, the degradation in performance was noticeable, but not dramatic. It is also interesting to note that because of the random draw of data files from the pool, it turned out that the set of three seed files were only part of the training data in 18 of the 200 runs, or 9%. These results would indicate that if one were to acquire a new important document that should start its own new group cluster, it may not be necessary to retrain the SOM.

In general, the results obtained by the bootstrapping runs agreed with those obtained in the initial SOM analysis. While it would be more desirable if the SOM mapped all files to the same classes at all times, in this study we were primarily concerned with the inflammatory writings. In this area the SOM did quite well. Most of the other articles covered a wide spectrum of topics. It is possible that many of the other articles would not normally be given to a system looking for phase changes.

Jackknifing

In Jackknife analysis, the statistics of the original data set are repeated with one sample left out of the set. That is, if a set contains N data samples, the statistics are computed N times with N-1 data samples each time. In each computation, a different sample is left out. To be consistent with the original SOM, 63 different realizations of the SOM were trained, using 62 files each. No training file was repeated in the same run. In all but 3 of the SOM realizations, all 3 seed files were present in the training set. Running the same analysis as described in the bootstrap runs above, the following results were obtained.

Table 3: Jackknife results outlining portion of runs SOM actually placed file with presumed group. Total number of runs=63.

Data file description	Presumed Group via manual and old SOM	Actual group(s) found by jackknife	Percent of runs in each group
Training data			
Ultimatum preceding attack	3	3	100
News about LTTE working to peace	1	1/other	98.4/1.6
Thimpu declaration (objectives of LTTE)	1	1/other/2	95/3/2
Article about peace	1	1	100
Article about killing of humanitarian workers	2	Other/2	65/35
Speech from LTTE leader (2004), inflammatory language	1	3/1	97/3
Speech from LTTE leader (1992) more conciliatory	1 or 3*	1/2/other/3	51/25/22.4/1.6
Testing data			
Article about 2004 tsunami	1	1/other/2	85.7/11.1/3.2
LTTE sympathizer	1	1/2/other	48/29/23
Blog accusing Sinhala Buddhists of racism	1	1/2/other	46/30/24
Tamil cuisine article	1	1/other/2	59/23.5/17.5
Denial of suicide attack by sympathizer	1	2/1	60/40

(All training data was selected from a possible set of 100 files. Thus, in the "training data" results in the table above, the file specified was in the training set, but may not have been selected by the algorithm for training purposes. For this reason, we would expect the algorithm not to obtain 100% correct classification for all files. In no case was the "testing data" viewed previously by the SOM.)

In the training set, with a few exceptions, the results seem to be consistent with those obtained from the bootstrapping techniques. In the case of inflammatory writings (group

3) 22 files from the training set of 100 were placed in this group over 90% of the time and 25 files over 50% of the time. Again, the results concurred with the previous runs.

The results with the testing data were unexpected. Essentially, the SOM was making a separate grouping for the test data a significant portion of the time. The precise reasons for why the behavior of the clustering algorithm differed so much from the bootstrapping runs are unknown. However, not a single file from the testing set was placed erroneously into group 3 by any of the 63 jackknifing runs.

Summary and Conclusions

This proof of concept study explored the relationship between textual data written by groups of interest and their relationship to subsequent group phase changes. The technique employed consisted of gathering writings from a group via the world-wide web, and extracting words and phrases from the text. The resulting text data was converted into a numerical matrix using MATLAB. The matrix was used to train a SOM, and clustering was performed on the resulting SOM mapping. Once the collection of web pages had been finalized, the process was entirely automated except at two points: the examination of the master phrase list for redundancies, and the selection of a small number of seed files for cluster groups.

Unlike other work in the area, we did not presuppose anything about the significance of words or phrases in the text. The master phrase list was generated automatically from the training set of files, though it was manually inspected for redundancies. The SOM learning algorithm was unsupervised, and aside from the selection of 3 seed files for clustering purposes, the class designation was also unsupervised.

After an exploratory search, the Tamil Tigers (LTTE) were selected to serve as a study group. The selection of the LTTE was based largely on the facts that they have exhibited phase changes in the past, that an archive of their writings existed, and that they wrote large amounts of text in English.

Data collection was conducted using a web crawling spider downloaded from the internet. Due to the quality of the spider, an inspection of the gathered documents was required to weed out obviously unrelated documents, such as articles about the Jewish holocaust.

Initial results demonstrated that the SOM did a good job of classifying inflammatory writings leading to violence vs. other negative verse, such as firm denials by the LTTE of acts attributed to them. The SOM found inflammatory writings by the LTTE to be the least linguistically diverse of the articles it analyzed.

In addition, the SOM seemed to do a good job of distinguishing inflammatory writings actually attributed to the LTTE from inflammatory writings penned by sympathizers maintaining web blogs. We did not seek to have this outcome in the beginning, but it was an interesting observation.

While a specific analysis of the inner workings of the SOM was not conducted, informal observation seemed to indicate that its classification was based more on sentiment, than individual word usage. For example, the number of times the word "kill" appeared in a document did not seem to have a direct relationship to how that document was classified.

A statistical analysis via bootstrapping and jackknifing was performed to add credibility to the findings. The statistical findings largely supported the initial SOM analysis. Namely, the SOM technique did well at classifying inflammatory text leading to terrorist attacks, articles with a more peaceful tilt, and inflammatory articles written by sympathizers and not the Tigers. In some cases, the bootstrapping and jackknifing pre-processing and SOM placed these documents in the correct group 100% of the time based on a total of 200 trials each. The SOM technique did less well on documents that could be considered questionable in classification by a human observer. For example, does an article condemning the killing of humanitarian workers belong in the group that addresses humanitarian issues, or in the inflammatory group? The algorithm placed it generally in the humanitarian group, though not with complete certainty.

The SOM did a good job of detecting behavioral changes in group activities. The availability of the archival record was inconsistent and sparse. We believe that with a more consistent paper trail, we could detect major attitude shifts, as well. If one were continuously monitoring the internet for writings from a group, and periodically retraining a SOM, the capacity to predict the larger shifts is a possibility.

From this proof of concept exercise we did uncover the ability to identify writings immediately preceding a terrorist attack. In fact, these writings were mapped to the *same identical cell* in the statistical analysis over 2/3 of the time. Such an observation could prove useful, as migration of data to this micro group on the map could serve as advance warning an attack is imminent.

One interesting data point we discovered in our analysis was that a document written by the LTTE signaling a move toward peace preceded an attack by GoSL troups. A worthwhile follow-on analysis would be a SOM based study of interplay between Tamil Tiger and Sri Lankan government writings. The Sri-Lankan government web sites are still up and running, and an operation to collect the necessary data would be straightforward.

This study sought to develop text analysis techniques to detect phase change in organizational writings. The results suggest that it may be possible that an unsupervised learning technique such as the SOM could be used not only to detect phase changes, but to provide an advanced warning of attacks, as well as classify whether a document is genuinely written by the group or by a sympathizer. All the results were obtained without editing the word and phrase list training data to emphasize sentiment or intent. If used to continuously monitor the writings of a group and re-trained periodically, the SOM could prove a useful tool in practice.

Why should a decision maker be interested in this type of tool? We have demonstrated the signal may be present in text that allows us to map a group writing to a subsequent violent action. Such analysis may help to support cause and effect relationships found among groups of interest. It can help provide leading indicators that precede events of interest. More importantly, if one wants to influence the actions of a group or alter an outcome, such a tool can help guide where and when one should make an investment.

At present, there are some limitations that we foresee with using adversarial text to predict adversarial actions. We will never be able to use this tool to definitively predict the actions of a group. We can only estimate the potential of a group to undergo a phase change or activity. But there are several improvements that we can make.

Any operational system will need a mechanism that collects on-line text in real time, determines its relevance, and passes the useful documents to the analysis tool. Useful data acquisition was a hurdle in this project in that separating appropriate, signal containing documents from the hundreds of documents that contained essentially noise had to be conducted manually. For example, (as was discussed above) a document comparing the Tamil plight to the Jewish holocaust caused the spider to begin collecting documents on the Jewish holocaust. Any future endeavors in system advancement should include development of a data acquisition and filtering tool to insure that only documents concerning the group in question are presented to the system.

In addition to acquiring quality data, there is also the issue of acquiring a quantity of data. Further analysis will need to be conducted to determine the amount and granularity of data that is required to provide a reasonable threat forecast.

A self-organizing map approach was selected for this study due to its long reputation as a useful pattern recognition tool. Other mathematical and pattern recognition techniques may be better suited for this application and should be examined as possible candidates in any future exploration.

Group dynamics occur across several dimensions. It is likely a dataset will contain several types of phase changes other than those of interest to operational personnel. Further work needs to be performed in the area to determine how well we can isolate one dimension of interest. In pursuit of this goal, it may be useful to examine other types of phase changes, such as the suggestion of future criminal activity (i.e., Enron email data set).

One limitation of the phase change system may involve the boundaries of the data itself. Although the internet itself is a relatively new medium, the use of only internet text to suggest phase changes is somewhat limiting. People do write and signal each other using newer and older media. Many conversations are not captured in text at all. The incorporation of network logs of mobile phone usage into the data set, for example, has the potential to strengthen greatly the system reliability to anticipate phase changes.

In the future, it would be useful to refine the technique to provide a graded scale of threat level instead of a binary indicator (danger/ no danger). The incorporation of a graded metric scale would help answer important questions such as how fast a group is approaching a phase change threshold. Further, metrics should be developed in a format such that the threat scale is useful to operational personnel and decision makers. Consultation with these persons to assess their requirements is essential for future advancement.

References:

[1] Doser, A., "Decision-aiding Algorithm for Interpreting Physiological Data in Hazardous Situations," *Proceedings of the Fifth IASTED International Conference on Signal and Image Processing*, August, 2003.

[2] Green, S., "Spin: Lexical Semantics, Transitivity, and the Identification of Implicit Sentiment," Ph.D. Dissertation, University of Maryland, 2007.

[3] Kohonen, T., *Self-Organizing Maps, 3rd edition*, (Berlin, Springer-Verlag, 2001).

[4] Oja, M., Sperber, G., Blomberg, J., Kaski, S., "Grouping and Visualizing Human Endogenous Retroviruses by Bootstrapping Median Self-Organizing Maps," *Proceedings of the 2004 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 2004.

[5] Shellman, S., "Process Matters: Conflict & Cooperation in Sequential Government-Dissident Interactions," *Security Studies* 15(4): 563-99, 2007.

[6] Shellman, S., "Machine Coding Nonstate Actors' Behavior in Intrastate Conflict," *Political Analysis* 16(4), 2008.

[7] Shellman, S., Hatfield, C., Mills, M., "Disaggregating Actors in Intrastate Conflict," forthcoming in *Journal of Peace Research* 47(1), 2010.

[8] Smith, A., "From Words to Action: Using Content Analysis to Explore the Psychological Dynamics of Groups that Engage in Terrorism," Ph.D. Dissertation, University of Michigan, 2003.

DISTRIBUTION

1	MS 1011	C. H. Duus, 6343
1	MS 1011	A. E. Speed, 6343
1	MS 1011	S. J. Verzi, 6343
1	MS 1011	C. E. Warrender, 6343
1	MS 1088	J. D. Morrow, 6341
1	MS 1088	J. S. Wagner, 6341
1	MS 1161	A. B. Doser, 5432
1	MS 1161	T. Stalker, 5432
1	MS-0899	Technical Library, 9536 (electronic copy)

Sandia National Laboratories