Quantifying Tactical Risk: A Framework for Statistical Classification Using MECH*

Xing Wang[1], Stephen George[2], Jason Lin[1], Jyh-Charn Liu[1]

[1]Texas A&M University, College Station, TX
{xingwang,senyalin,liu}@cse.tamu.edu
[2]U.S. Department of Defense, Washington, D.C.
ticom.dev@gmail.com

Abstract. This paper presents a statistical classification framework for classification and prediction of asymmetric conflict (AC) locations. Various methods of data normalization and feature reduction are paired with supervised machine learning training algorithms to train classifiers. A set of 77 features derived from the MECH Model (Monitor, Emplacement, and Command/Control in a Halo) were used to train the classifiers. The framework has been implemented and tested on real-world improvised explosive device and direct fire data collected from the conflict in Afghanistan in 2011-2012. In testing, the classifiers achieve high accuracy, with human behavior-related features (visibility and population) exhibiting more significant statistical differences. The performance is found to be insensitive to the type of training algorithm. Accuracy is positively correlated to the training data size as expected, but we could achieve rather good performance using quite few data. Cross-region training and prediction shows that the risk classifiers are region-specific, which also might reveal the tactics differences across regions.

Keywords: Statistical pattern, Risk averse behavior, Asymmetric conflicts, Machine learning, Afghanistan

1 Introduction

Asymmetric conflict (AC) often pits a weaker attacker against a stronger target. To increase the probability of success, the attacker carefully assesses and optimizes attack-related tactical decisions. In [1], we propose a tactical behavior model for ACs called MECH that captures these decisions as features associated with common AC roles. A *Halo*, or annulus, surrounding the *Emplacement* site defines locations available for *Monitor* and *Control* activities. At core of MECH is a set of 77 features that describe the local terrain, social/cultural features associated with nearby population centers, and AC actor positioning, under the assumption of a risk-averse attacker.

*This work was supported in part by an ONR grant N00014-12-1-0531 and a National Defense Science and Engineering Graduate (NDSEG) fellowship. Opinions, findings and conclusions or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsors. Corresponding author is J.C. Liu.

SBP 2015, p. 446, 2015.

In this work, we use supervised statistical machine learning (SML) techniques to predict high-threat locations for improvised explosive device (IED) and direct fire (DF) attacks. Raw measurements are treated as features and processed with normalization and reduction for characteristic exploration. Stepwise feature selection (STP) and Principal Component Analysis (PCA) techniques are used for feature reduction. k-Nearest Neighbors (KNN), discriminant Analysis (DA) and Support Vector Machines (SVM) algorithms train a binary (attack vs. no-attack) classifier. Using a dataset of roadside attacks in Afghanistan from 2011-2012, together with randomly selected non-incident locations, we develop a system for prediction of roadside attack locations that incorporates feature extraction and classifier training.

Overall, the classifiers demonstrate good generalization ability on the tested data set. Experimental results show that features correlated with risk-aversion and local population characteristics offer high discriminant ability in comparison with non-specific geomorphometry features. Experiments on cross-region training and evaluation show that the classifiers are region dependent. Although performance is influenced by the size of the training data set, but we can get nearly optimal performance with only few data.

The rest of the paper is organized as follows: Related Work in Section 2 is followed by Section 3 that presents features and their tactical meaning. Section 4 introduces the SML framework for road risk prediction and Section 5 presents analysis of experiments designed and executed using the framework.

2 Related Work

Deitchman [2] modified the conventional Lanchester warfare model to describe guerrilla warfare, where attacking power is a key parameter of the model. In AC, the attacking power of a team is a factor of both the quality of their weapons and the protection of the team. This duality is seen in agent-based guerrilla war [3], which analyzes the relationship between mobility and the rate of success. However, describing real world terrain in terms of mobility is a difficult problem. Perry and Gordon note that more effective and reactive techniques and procedures to describe counterinsurgencies are necessary [4]. Also, the need to tailor and update adversarial models quickly requires an unrealistic degree of situational awareness and data availability.

Explicit behavior modeling requires access to detailed decision-making paradigms. An alternative modeling approach uses SML to capture the patterns of past events in optimized classifiers that can be used for prediction of future situations. The development of statistical theory and advancement in toolkits has made it easier for an operator without advanced knowledge to conduct statistical analysis to achieve analytics goals. Bernica [5] tried to correlate political, geographic, social, economic and cultural factors with successful insurgent groups, in order to take preemptive actions before a group becomes too strong. Breiger [3] used logistic regression to propose a relationship between organization/environment features and the pursuit/use of chemical, biological, radioactive or nuclear weapons (CBRN). They also used eigenvectors to profile potential CBRN terrorist groups. [6] uses linear discriminant analysis to explore factors affecting the likelihood for a person to become insurgent.

3 Road Risk Assessment Based on SML

We formalize the task as a binary classification problem, classifying each location as useful or unfit for attack. Binary classification is the outcome of the linear process shown in Figure 1 that includes feature extraction, data selection, feature normalization and reduction, classifier training and evaluation.

First, each potential AC event location is represented by 77 features using the MECH model [1] that capture the locational relationship of attackers and targets in an AC event. Relative visibility is a key component in the qualitative description of tactics. Viewshed and sparse viewshed derived from digital elevation maps (DEM) are key structures that store all information about intervisibility between *Emplacement* locations and the neighboring region defined by the *Halo*. The first three groups of features in Table 1 are proposed based on tactical analysis. Additionally, general features based on the population and simple geomorphometric analysis are included. Given that available historical data does not contained full details of particular roadside attacks, it is impractical to obtain optimal range parameters for each feature. Instead, on the basis of MECH behavior model, we assume that attackers and targets operate at different ranges from the roadside attack location. Each feature type is expanded into different ranges (shown in the title of each group of features) and each range is treated as a separate feature. The proposed feature set creates a mining grid for the SML learning algorithms to capture hidden patterns.

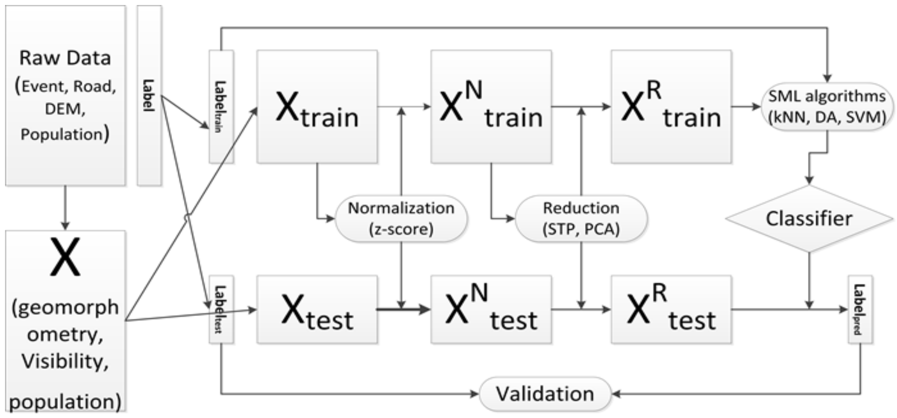

Fig. 1. Workflow of SML for risk assessment of potential AC event locations.

After feature extraction, the data could be represented as a $n \times 77$ matrix X with each row representing one location and a column label vector $l \in \{0,1\}^n$. l_i is the label of the ith location with 1/0 indicating the location with/without event history. Then we need to randomly down sample the non-event road location and split the data into train part X_{train}, l_{train} and test part X_{test}, l_{test}. The downsample of non-event locations is due to the unbalance between the event and non-event data which will result in classifiers in favor of majority type (non-event). Besides, we leave out test data to evaluate the performance of classifier as supervised machine learning algorithm might over fit the raw data. In this work, we order the event samples based on date and take the first 2/3 of the samples as training data the left as test data.

Before feeding the feature matrix X_{train} into the SML algorithm, we need further feature normalization and reduction to make them more suitable for the algorithm.

Table 1. Features and relevant behavior

G_1: Viewshed from victim's view (window radius 100-350, 350, 500, 1K meters)	
Visibility Index (4)	The more region the defender can see, the more safe it is.
Shape Complexity (4)	The more complex the visible region, the more difficult to defend.
G_2: Sparese viewshed from vicim's view, # of directions (4,8,16,32,64)	
Local openness (5)	The openness of the region near emplacement.
Distance to invisible region (min/mean/max) (15)	The nearest invisible region could be treated as cover for attacker, the farthest invisible region could be good location for Monitor
Planimetric area (5)	Sparse viewshed version of visibility index.
Rugosity (5)	Sparse viewshed version of roughness within the visible range
Shape Complexity 3D (5)	Shape complexity based on sparse viewshed.
G_3: From Monitor/Control view, route range (100, 250, 500, 1K, 3K meters)	
Route visibility near emplacement (min/median/max) (15)	Evaluate observability over route from possible locations for attackers.
G_4: General Terrain Features (Window radius 50, 100, 350, 500,1000 meters)	
Elevation (1), Slope(1), Convexity(1) and Texture(1) Elevation range (5), Elevation roughness(5)	
G_5: Population related features, population threshold (1, 1K, 10K, 50K, 100K)	
Minimum distance to city of at least certain size (5)	Populated areas act as support for attacks.

Features with large scale tend to play more important role in SML algorithms that use concept such as distance (kNN) and margin (SVM), while they might not be the best to provide discriminant ability. In this work, we adopt z-score normalization to overcome this problem. For jth feature, we get mean μ_j and standard derivation σ_j based on the training data. Then for each row x in X_{train}or X_{test}, the normalized jth feature value is $x_j^n = (x_j - \mu_j)/\sigma_j$. We conduct feature reduction to remove noise, remove correlation among features, and reduce dimension due to the curse of dimensionality. Two dimension reduction methods are used here. Unsupervised principle component analysis (PCA) [7] transforms the data from high dimensional space into low dimensional orthogonal space which conserve most of its variance. Supervised stepwise feature selection [8] is a regression-based iterative greedy algorithm. It evaluates the importance of the feature based on coefficients of the linear regression model. When the coefficient for a feature is near 0, this feature is not contributing to the regression.

After the feature reduction we get $n \times p'$ data matrix X^R, and together with the label vector l they will feed the SML algorithm. We use three kinds of algorithms based on different heuristics [8].

- KNN method is based on the heuristic of density estimation. The density function for each class at each location in high dimensional feature space is estimated by the number of instances of current class in unit space volume around the current location. The label of a new instance is assigned to be the class with largest density value at the location of this new instance. This method is sensitive to rescale of the feature space, and density estimation is not accurate at high dimensionality.
- Discriminant analysis finds a projection that maximizes between-class variance and minimizes within-class variance. This method assumes that each class has a Gaussian distribution and that mean value is the main difference between different classes. The assumption is questionable for this problem.
- Support vector machine performs structural risk minimization. This theory shows that an algorithm can achieve the minimal risk of the linear model by maximizing the margin -the minimum distance of an example to the decision hyper plane. This method overcomes the curse of high dimension, and the kernel trick can map features into high dimensional space that is more separable.

4 Experiment, Result and Analysis

In this work, we applied our framework to an AC event dataset containing IED and DF attacks collected across Afghanistan between February 1, 2011 and August 23, 2012. (Two periods are missing: 09/12/2011-11/8/2011 and 02/14/2012-03/31/2012.) The performance and comparative analysis of classifiers trained on the whole data set using various feature reduction methods and machine learning algorithms is presented. Feature rank is also presented to show their relative contributions. We also explore the location sensitivity of the classifiers by training classifier for one command region in Afghanistan and applying it to another region.

4.1 Data Preparation

This analysis is constrained to roads and events along roads. Road data is collected and maintained by the Afghanistan Information Management Service. We discretize continuous roads to be a set R of locations according to the digital elevation map (DEM) from the ASTER Global Digital Elevation Model Version 2. DEM offers digital elevations with a horizontal resolution of approximately 30 meters.

The asymmetric warfare events in our experiment are from the ISAF-NATO Civilian Integration Team. This dataset consists of a variety of events but we only focus on IED and DF in this work. The complete set of insurgent attacks are recorded as transactions with location, type and time stamp. Let T be the set of attack locations. But we cannot use all such locations as some events happen away from known roads. Thus, we filter out event locations that are more than 100 meters from known roads and get a final event set *E* as the event class. For the non-event class NE, we assemble a set of discrete road locations known to be at least 250 meters from a E locations. This compensates for suspected positional and estimation errors in the event data. The resulting datasets include T_{IED}=13,295 events, T_{DF}=16,609 events and NE=3,313,984.

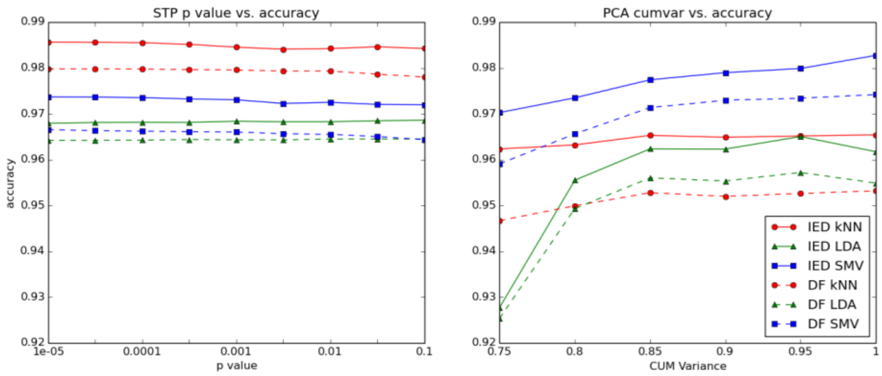

Fig. 2. Effect of feature reduction parameter on prediction accuracy

4.2 The Effect of Parameters of Feature Reduction on the Performance

Fig 2 shows the trend of accuracy change when the parameters of feature reduction methods are adjusted. We find that STP is less sensitive to parameters compared with PCA. Overall, STP performs better than PCA methods. As the threshold for *p*-value decrease, we achieve little improvement on performance. For PCA, as the cumulative variance threshold decreases, performance decreases significantly. This means that dimensions with little variance also contain information and should not be removed. The large decrease is possibly because PCA is designed for signal processing and might not fit multi-modality fused features.

4.3 Performance of Different Reduction Methods and ML Algorithms

We use machine learning algorithm with default parameters from the sklearn[1] package. The cumulative variance threshold for PCA is 1 and *p*-value threshold for STP is 10^{-5}. To validate stability, the experiment is repeated 10 times. Mean accuracy is reported in Table 2. All standard deviations are less than 0.0015.

In general, there is no significant difference for different machine learning configuration. No machine learning method is consistently better than others with respect to

---

1 http://scikit-learn.org/stable/

feature reduction methods. The best configuration for both the IED and DF event prediction is no feature reduction and SVM.

Table 2. Accuracy for IED/DF risk prediction using kNN, DA and SVM

	IED			DF		
	kNN	DA	SVM	kNN	DA	SVM
None	0.9696	0.9688	**0.9876**	**0.9784**	0.9644	0.9582
STP	**0.9856**	0.9679	0.9737	**0.9798**	0.9642	0.9666
PCA	0.9654	0.9617	**0.9828**	0.9532	0.9549	**0.9742**

From the view of feature reduction method, none is consistently better than the others with respect to each machine learning methods. For similarity-based method kNN, STP gives the best performance, likely because it gives more weight to features that best separate the two classes. For DA, the accuracy mostly unchanging, primarily because it is insensitive to scale and linear feature transformations.

4.4 Features Comparison

There are two ways to evaluate the contribution of features on the final performance. One way is to use each feature or each group of features to train classifier and the performance is shown in Fig.3. Only the route visibility features could result in very good performance. Population features also have good performance. In comparison feature group $G_1$and G_2 based on victim's view are not very helpful.

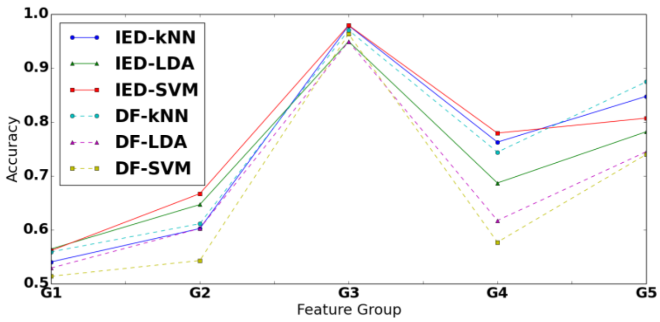

Fig. 3. Performance of each feature group

Another way is the ranking based on the weight in feature reduction as shown in Table 3. The STP method is uses a supervised feature reduction technique and the resulting feature ranks are in accordance with the performance in Fig.3. Interestingly, more weight is placed on features constrained by human factors. The route visibility feature identifies locations where a *Monitor* or *Control* actor would have a longer than normal time period where the target was continuously visible, allowing more attack preparations to occur. Visibility index and shape complexity based on viewshed also rank highly. The window size is larger -500 and 1000 meters, meaning that the target will be exposed to attack over a greater distance before finding safety out of sight of the attacker. Finally, no feature based on sparse viewshed is ranked in the top 10. This might mean the sparse viewshed analysis process loses too much information.

PCA is an unsupervised dimensionality reduction method that gives more weight to features with larger variance, such that noise is removed. The rank of features for PCA is based on the absolute value of coefficients for the first new dimension which has the largest variance. A high proportion of the weight is drawn from features based on sparse viewshed such as radius and planimetric area. Other viewshed features like visibility index and shape complexity index fail to show any strong discriminant ability. This may indicate that PCA is not the best option in above analysis.

Table 3. Feature rank by two feature reduction methods for IED and DF

IED-PCA 1	IED-STP	DF-PCA 1	DF-STP
meanrad32	mincea100routevis	meanrad32	mincea100routevis
meanrad16 mean-	maxcea100routevis	meanrad16	maxcea100routevis
rad64 meanrad8	scid1000	meanrad64	scid1000
visidx500	dist100kpeople vi-	meanrad8	visidx1000
scid100-350	sidx1000	scid100-350	dist100kpeople
meanrad4 vi-	dist50kpeople	meanrad4 vi-	rng100
sidx100-350 vi-	scid500	sidx500 vi-	rgh100
sidx350	visidx500	sidx100-350	visidx500
scid500	maxcea1000routevis	visidx350	dist50kpeople max-
	maxcea250routevis	scid500	cea1000routevis

4.5 Performance for Training Set of Different Sizes

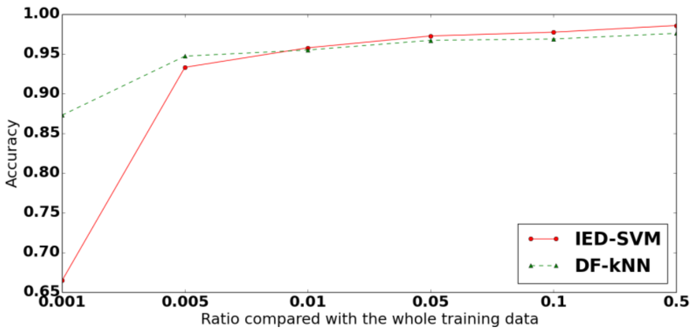

Fig. 4. Training data size vs. prediction performance

Supervised methods are sensitive to the quality and quantity of the training dataset. Knowing the sufficient size of training data can reduce the effort of collecting data. Fig. 4 shows that there is a positive correlation between the size and the accuracy. Before ratio 0.005 (about 100 samples), there is a huge increase in the performance and increase very little afterward. This ability to train a usable classifier with few data can help make quick reaction in a new environment.

4.6 Generalization Ability Across Different Regions

Besides size of training data, another important issue is whether classifier trained in one region could be used in another region. For each command region[2], we split the event set into train and test sets using a 2:1 ratio. For other regions, we just use the whole dataset. Then we randomly selected a corresponding number of NE locations from the region. The classifier is trained using the training data of current region. Then prediction is made on the test data of current region and the data of other regions. The process is repeated 10 times for each region, and results for IED and DF risk prediction are reported in Table 4. The standard derivation of the experiments are mostly less than 0.01. Each row uses the same classifier and each column uses the same test set. The prediction performance is usually better when using classifiers of the same region, except for DF risk prediction on region with less data. Compared with IED, the DF prediction performance is more region specific. This could be caused by the difference in commander training or different terrain constraints.

Table 4. Prediction accuracy based classifier of other regions

IED					
region(#of evt.)	e	n	s	sw	w
e(3084)	**0.9820**	0.9762	0.9728	0.9565	0.9456
n(353)	0.9377	**0.9648**	0.9461	0.9331	0.9260
s(4513)	0.9637	0.9773	**0.9818**	0.9654	0.9755
sw(4791)	0.8729	0.8858	0.9470	**0.9888**	0.9501
w(554)	0.9402	0.9459	0.9504	0.9429	**0.9789**
DF					
region(#of evt.)	e	n	s	sw	w
e(5841)	**0.9767**	0.9741	0.9505	0.9435	0.9204
n(193)	0.8634	**0.9338**	**0.9604**	**0.9530**	0.9334
s(3912)	0.8848	0.9648	**0.9803**	0.9769	0.9497
sw(6115)	0.7975	0.8640	0.9570	**0.9897**	0.9305
w(548)	0.9132	0.9370	0.9492	**0.9689**	**0.9669**

4.7 Conclusion

In this paper, we present a general pattern mining system based on a set of features derived from a tactical behavior model called MECH and some general demographic and geographic data. Experimental results show the tested classifiers, together with their preprocessing steps, have good generalization abilities as defined in the ML community. The accuracy does not change a lot for different feature reduction and machine learning methods. But the features have huge impact on the final perfor-

---

2 http://www.nato.int/ISAF/structure/regional_command/index.html

mance. Human behavior related features (route visibility and population) are very important, which is consistent with the weight from STP feature selection. Follow up study on analysis of specific error cases can be easily performed as needed. We further conduct more experiments about the training data set. Region and size of the training data can affect the performance a lot. The region-specific property of the classifier might also reveal the tactics difference across regions. This might increase the difficulty in building classifier, but it could be mitigated by the fact we could build nearly optimal classifier with very few data.

For future work, we are in the process of implementing a software system to automate the training and evaluation steps, in order to support a broader range of experiments as well as cross analysis with the behavior-based simulation reported in [9].

5 References

[1] S. George, X. Wang, and J.-C. Liu, "MECH: A Model for Predictive Analysis of Human Choices in Asymmetric Conflicts," presented at the Submitted to SBP 2015, Washington D.C., 2014.

[2] S. J. Deitchman, "A lanchester model of guerrilla warfare," *Operations Research*, vol. 10, no. 6, pp. 818–827, Nov. 1962.

[3] R. L. Breiger, G. A. Ackerman, V. Asal, D. Melamed, H. B. Milward, R. K. Rethemeyer, and E. Schoon, "Application of a Profile Similarity Methodology for Identifying Terrorist Groups That Use or Pursue CBRN Weapons," in *Social Computing, Behavioral-Cultural Modeling and Prediction*, J. Salerno, S. J. Yang, D. Nau, and S.-K. Chai, Eds. Springer Berlin Heidelberg, 2011, pp. 26–33.

[4] W. L. Perry and J. Gordon, "Analytic Support to Intelligence in Counterinsurgencies," RAND National Defense Research Institute, 2008.

[5] T. W. Bernica, V. E. Guarino, A. J. Han, L. F. Hennet, M. A. Mitchell, M. S. Gerber, and D. E. Brown, "Analysis and prediction of insurgent influence for U.S. military strategy," in *2013 IEEE Systems and Information Engineering Design Symposium (SIEDS)*, 2013, pp. 161–166.

[6] J. Doran, "Iruba: An Agent-Based Model of Guerrilla War Process," in *In Representing Social Reality, Pre-proceedings of the Third Conference of the European Social Simulation Association (ESSA*, 2005, p. 205.

[7] I. Jolliffe, "Principal Component Analysis," in *Encyclopedia of Statistics in Behavioral Science*, John Wiley & Sons, Ltd, 2005.

[8] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Verlag, 2001.

[9] J. Lin, B. Qu, X. Wang, S. George, and J.-C. Liu, "Risk Management in Asymmetric Conflict: Using Predictive Route Reconnaissance to Assess and Mitigate Threats," presented at the Under review for SBP 2015, Washington D.C.