Probabilistic Prognosis of Societal Political Violence by Stochastic Simulation

Using Principal Component Analysis and Support Vector Machines

André Brahmann, Uwe Chalupka, Hendrik Rothe, Torsten Albrecht
Institute of Automation Technology,
Measurement and Information Technology
Helmut-Schmidt-University
Hamburg, Germany
Andre.Brahmann@hsu-hh.de, Uwe.Chalupka@hsu-hh.de, Hendrik.Rothe@hsu-hh.de, albrecto@hsu-hh.de

***Abstract*—The current paper deals with the probabilistic prognosis of societal political violence levels of countries in the context of crisis prevention. The baseline is formed by two freely available datasets, whose relevance gets clarified in the first part of the paper. From these, a classification and a prediction modeling problem can be derived for which a Principal Component Analysis together with Support Vector Machines (SVMs) can be utilized as useful methods. Different SVM kernel functions have been investigated. To further perform the prediction step, a statistic modeling approach has been chosen that includes the computation of occurrence probabilities by stochastic simulation.**

Keywords-stochastic simulation; Major Episodes of Political Violence; principal component analysis; classification; support vector machines.

I. INTRODUCTION

Within the United Nations, the European Union and the North Atlantic Treaty Organization (NATO) crisis management, conflict prevention, security of trade routes, humanitarian aid, reconstruction activities as well as international cooperation are some of the main foci nowadays [1][2]. Thus, it is of high value to determine and anticipate possible geographic regions of interest in time [1]. These kind of regions will be defined by the term “hotspots” in the current work and the question arises how to identify and predict them.

A good clue to determine possible hotspots is given by the *Major Episodes of Political Violence* (MEPV) dataset, which is provided by the *Center of Systemic Peace* [10]. In the context of the current work, it is used as reference to characterize hotspot regions, because of its special properties. It summarizes the various occurrences of violence due to intrastate (civil, ethnic) and interstate conflicts on country level and transforms them into normed warfare magnitudes as time series on a yearly basis [10]. Thus, it allows a continuous assessment and objective comparison for each year with the warfare magnitude being proportional to the hotspot level. Other datasets, similar with regards to content, list singular events instead, which makes it hard to perform an continuous analysis on a yearly basis.

Although the usage of this dataset leads to a very convenient way of determining possible hotspots, it causes two distinct issues. The first one is the lack of data for certain regions, so that it is impossible to achieve a sufficient covering of the entire world. The second issue is the unsatisfying feasibility to forecast upcoming hotspots due to insufficient information contained in the MEPV-data for this task.

For that reason, governance indicators provided by the *World Bank* [11] comprising hundreds of underlying societal political factors on a yearly basis are additionally used. These indicators include direct and indirect measures of violence [3]. This makes the dataset basically also applicable within the given scope of hotspot assessment. Thus, the question may arise why not to use these governance indicators as a complete substitution for the MEPV-dataset instead. The reason is, the governance indicators comprise a lot more information and thus span more than one dimension (see Section II.B). This introduces the problem, that one has to decide for which values in what combination a hotspot can be assumed. So, statistical methods and data mining techniques are applied to recognize patterns that are related to certain values in the MEPV-dataset. With this additional information, it is possible to overcome the two issues mentioned above.

In the present paper, the basic approach to connect the level of violence to a certain tuple of indicators is shown and examined to reconstruct missing MEPV-data. Furthermore, a stochastic approach to predict the development of societal political violence by stochastic simulation is presented. Combining the reconstruction and prediction methods, finally it will be shown how MEPV-data can be predicted including the declaration of probabilities. Basically, the applied methodologies are Principal Component Analysis (PCA) and Support Vector Machines (SVM). In the first part of the paper, these methods will be briefly summarized, their relevance clarified and finally explained in which way they are applied for the current problem.

While the presented forecast modeling itself is yet simple, it reflects the underlying methodology, which can be transferred to any higher quality forecasting model. The regarded time frame of the prognosis is targeted for a short term basis of up to five years.

In the next section, the used datasets will be described in more detail. Afterwards, it is explained how and why (at a first glance) a PCA is applied to the current problem. Then, the data classification process and its relevance will be ex-

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-371-1

plained, comprised by a motivation and description for the use of SVMs. Finally, the stochastic simulation model will be presented and linked with the PCA and SVM methods. Due to the dependence of the SVM classification quality to the used kernel function [12], different kernels will be investigated for the described datasets.

II. DESCRIPTION OF USED DATA- AND TOOLSETS

A. MEPV-Data

The dataset is provided for free on an annual basis and lists cross-national, time-series data on interstate, societal, and communal warfare magnitude scores (independence, interstate, ethnic, and civil; violence and warfare). The value of interest here is the civil violence (“CIVTOT”-value), also later referred to as “MEPV-level”. It is coded from zero to ten using a discrete scale, where zero means no violence and ten means extermination/annihilation [10]. In general, low levels of violence are smaller than four [9]. Currently values are available from 1946 to 2012 for a total of 167 countries [10].

B. Worldwide Governance Indicators

This dataset is distributed by the *World Bank* and comprises a tuple of six indicators as times-series for 215 countries: (1) Control of Corruption, (2) Government Effectiveness, (3) Political Stability, (4) Regulatory Quality, (5) Rule of Law and (6) Voice & Accountability. Each indicator ranges approximately from -2.5 (bad) to 2.5 (good). The most recent set covers a time span from 2002 to 2012 on a yearly basis. The indicators can also be obtained for free, see [11].

C. Toolset

Before any calculation and examination could be done, some decisions regarding the need of specials tools have to be made. For prototyping and for a later verification, all algorithms were implemented in *Matlab R2014a* [14]. It is of great value that *Matlab R2014a* provides convenient implementations of the PCA and SVMs in its statistics toolbox. The first prototype relies exclusively on these two modules. The rest is done using standard functions provided by *Matlab* itself.

Beside the convenience of a *Matlab* implementation, there are also some downsides. First of all, the SVM implementation in *Matlab* does not provide native multiclass classification. Secondly, the creation of a custom kernel modification turned out to be difficult, as not being well documented. Furthermore, it becomes very tedious to maintain the code if a project exceeds a certain size. For these reasons, the working algorithm was implemented in the programming language *C#*. Due to its object oriented paradigm, it is very well suited for projects dealing with structured data and provides a very convenient syntax. While there is no native implementation of an SVM or PCA available in this language, as a downside, mature third party frameworks are available however. For the current application, the *Accord.Net* framework [13] has been chosen. This framework provides a large variety of statistical modules, including multiclass SVM and PCA. The framework is available for free and Open-source. Furthermore, it provides a very useful interface, which allows a convenient way to create any type of a custom SVM kernel.

So, every calculation described in this paper could be reproduced using either *Matlab* or *C#* in conjunction with the *Accord.Net* framework.

III. APPLICATION OF PRINCIPAL COMPONENT ANALYSIS AND TRANSFORMATION

A. Dimension Analysis

As a first step, the dimensionality of the information contained in the six governance indicators is verified. Therefore, a PCA as described in [1] is applied.

The PCA is basically an approach to transform a tuple of observations, which are potentially correlated, into a set of uncorrelated variables. The result of this transformation, which is essentially a translation and a rotation, are the principal components (PCs) of the given observations. For a useful application of the PCA, the data has to be centered and standardized in a preceding step. Furthermore, to ensure non-corrupt data containing no outliers, an outlier test like the *Wilks*'s test [7] has to be performed.

The next step is the investigation of a possible reduction of the dimension of space. For that stage, the values of all eigenvalues of the correlation matrix have to be examined. The higher a value, the higher is the variance of the related principal component and therefore, the importance for declaring the features of the entire dataset. In Figure 1, the sorted percentages of all eigenvalues are shown. As one can clearly see, the first three principal components explain more than 96% of all observed features in the dataset.

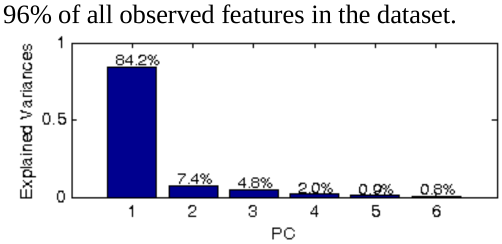

Figure 1. Percentage of all eigenvalues of the correlation matrix

As a first guess, it seems obvious to neglect the three inferior components. To ensure the correctness of this approach, a significance test with an error probability of 1 % on all the eigenvalues has been performed. Finally, this test certifies every eigenvalue as significant, so that a safe dimension reduction cannot be guaranteed.

B. Determining the Principal Components

Though proven with certain significance that there should be no dimension reduction performed, the results of the PCA are further used to compute the PCs of the indicators. The reason for that is their application in the data prediction step, as explained later.

For this, the eigenvectors of the previously calculated correlation matrix are computed. The eigenvectors should be sorted and scaled by the square root of their corresponding eigenvalue. When summarized into a matrix, a transformation matrix M is gained. The final transformation operation then is given by equation (1):

$$PC = Data \cdot M. \tag{1}$$

IV. DATA CLASSIFICATION

A. Methodology

As a next step, it is required to develop a systematic scheme that allows to determine to which MEPV-level a certain indicator tuple for a single year and country corresponds. In Figure 2, scatterplots for each pair of the first three PCs are given as example, comprising the corresponding color coded classes in terms of MEPV-level.

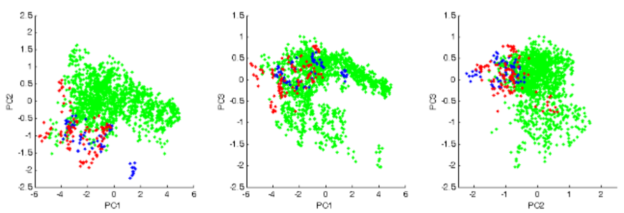

Figure 2. Scatterplots of the first three principal components (left: PC1 & PC2; middle: PC1 & PC3; right: PC2 & PC3) with the corresponding color coded classes: green represents no violence, blue represents a raised violence level and red represents hotspots.

As it can be seen the data seems to be spatially separable by adopting some classification method. In the next section, we take a closer look on possible classifications methods in the context of the given data.

B. Choosing the Classifier

As the simplest solution for solving the classification problem, the PCA could be used to discriminate the dataset. The drawback of using this method is that it can only be used for the classification of linear separable classes. A closer look at Figure 2 discloses that the underlying problem is not linear, therefore the PCA provides only a poor classification quality.

Fortunately, there are currently more advanced methods available. Discriminant analysis [16], neuronal networks (NNs) [15], and SVMs [6] are also known for solving classification problems.

With the discriminant analysis as a multivariate, statistical approach it is possible to classify via a linear, quadratic, *Bayesian* and *Mahalanobis* distance discriminant function. As with the statement above, a linear discrimination function is not appropriate. Choosing the *Mahalanobis* distance as classifier, one gets the best results for this particular problem. The classification error in this case is about 12 %. Although this means that we are getting correct answers in approximately 88 % of all cases, it is worth examining a more refined approach for that particular classification problem.

NNs might be a promising choice as being able to solve non-linear and also complex classification problems like pattern recognition in image processing. Nevertheless, they suffer from their theoretical limitations. For instance, when using back-propagation as learning method, they usually converge to locally optimal solutions. In this particular case, SVMs can offer a major improvement [5]. On top of that, by picking support vectors, SVMs choose the model size on their own [5] and they are developed using a sound theory first [5]. The lucid theoretical foundation of SVMs stands in contrast to the theoretical base of NNs.

SVMs are well explained in [6] and the following description uses this as a foundation. As the starting point for an SVM classification, we need a quantity of training objects whose correct classification is known (this is also true for the discriminant analysis mentioned earlier). An object could be regarded as a vector in a vector space. The SVM tries to fit a hyperplane into the given vector space, which separates the given classes. At the same time, the distance between the nearest vectors and the hyperplane is maximized and these vectors are called the support vectors.

It is apparent that a hyperplane is capable of discriminating linear separable vectors only. At this point, the SVM uses the fact that nonlinear separable vectors become linear separable in higher dimensional space. Therefore, a transformation into higher dimensions has to be performed and the hyperplane could be fitted into the training data set. After that process, the hyperplane must be transformed into the original space, which leads to a nonlinear hypersurface, which could be also non-contiguous.

This algorithm causes one serious issue, namely, the transformation into higher dimensions is computational expensive; therefore, an alternative approach was invented. A method called the "kernel trick" solves this issue. The trick is to use an appropriate kernel function, which describes the hyperplane in higher dimensions. If applied, the for- and backward transformation into and from higher dimensional space can be achieved without computing it directly. Therefore, the art of using SVMs is to choose the correct kernel function, which is well suited for the underlying problem. That discussion will be postponed to Section VI.

With a well suited kernel function, it is possible to distinguish between all MEPV-levels with no error, representing a perfect classification. Anyway, a nearly perfect classification of the training data might cause the problem of an over-classification. This means that the trained classifier gets too specific and hence sensitive to minor changes. This might lead to too many misclassifications when performing the prediction step (this problem also applies to NNs).

So far, the discussion neglected the fact that the underlying problem is a multi-classification problem. The traditional approach is only suited to distinguish between two classes. To overcome this issue, the classification between c classes has to be subdivided into c classifications of two classes. Although this was not explicitly pointed out, it was taken into account, so that the given errors are valid for the generalized classification case of all MEPV-classes.

V. DATA PREDICTION VIA STOCHASTIC SIMULATION

A. Approach

As it is intended to perform a prognosis of the societal, political level of violence, a prediction of the development of the indicator data is required. Of course, there is no way to

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-371-1

achieve this with reliable results, as the social political processes in countries are too complex.

The most simple prediction approach would be an extrapolation of the indicator data based upon a linear regression model not regarding side effects. Such could be spillover effects between bordering countries with similar cultural and political backgrounds as most recently observed during the *Arab Spring*. A simple extrapolation does not imply the most likely development of the indicators. To consider probabilities, the actually chosen approach is to perform a statistic modelling of the indicator development based upon observed values.

B. Statistic Modelling

The first step is to derive the indicators for a specific country by time to obtain their annual slopes. From the distribution of the slopes the likelihood can be obtained for which an indicator will develop into a certain range of directions by integrating the probability density function (PDF) of the underlying distribution. At the other hand, it is possible to compute ranges of slopes for given probabilities. This reflects the approach done for the current application. To formally cover the whole probability range, the underlying PDF is split by n quantiles into equally sized ranges of the probabilities $1/n$. For each quantile range, the corresponding n mean slopes are computed from the PDF. In other words, they represent the n cases how an indicator might develop with identical probabilities. Finally, this step is performed for all six indicators. The predicted indicator values are then constructed for a certain year by using their last known values and the computed slopes, resulting in a set of six by n-indicator-tuples. From these, a total of p possible cases can be constructed on how all indicators might develop. This results from a permutation of each computed possible development of each indicator with

$$p = n^6. \tag{2}$$

One yet unconsidered issue is that the indicators develop partially dependent on each other. The performed permutation however is only valid if the indicators would be independent from each other. This condition can be met if the PCs of the indicators are used instead of the pure, untransformed ones. All the previously described steps remain the same for the usage of the PCs instead of the original data.

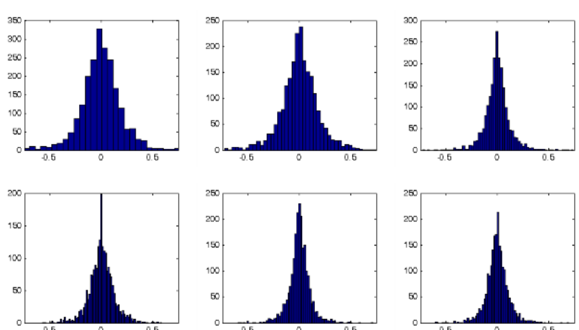

Figure 3. Histogram of the slopes of all PC values (first PC at upper left, sixth PC at lower right)

A remaining question is the type of PDF that is applicable to the current problem. For this, the complete distribution of the slopes for each PC of all countries have been investigated. It has been proven that the underlying PDF can be assumed to be a normal distribution. In Figure 3, the histograms of the slopes of the six PCs are given to exemplify that. The required sigma-value is obtained separately for each country by computing the standard deviation of the slopes of each PC for all years of a current country.

C. Classification

So far, p occurrences of indicator tuples per predicted year and country have been computed. Finally, of interest is the probable MEPV-level. Therefore, the classification methodology as described before by using SVMs can be applied. This requires that a SVM with an appropriate kernel is trained in advance by using a time span where both, indicator data and reference MEPV-data are available. The application of trained SVM to the predicted indicator tuples yields p MEPV-level occurrences per predicted year and country. The selection of a proper SVM kernel is addressed later in the next section.

As indicated by equation (2), the computational effort for classification dramatically increases if smaller sized quantile ranges are used. While for tertiles (spanning three ranges with each covering 33 % probability) 729 permutations occur, quintiles for example already produce 15625 permutations. These numbers then equal the count of indicator tuples which need to be classified for every predicted year and country.

Beside the option to avoid too small quantile ranges it is also possible not to include the less significant PCs (e.g., the 5th and 6th PC, see Figure 1) in the permutation process. Thereby, another advantage of using the PCs instead of the untransformed indicator data for classification is given.

TABLE I. MEPV PREDICTION, EXAMPLE OF CHAD

Year	MEPV-Level Occurrences (Σ = 729)					
	0	1	2	3	5	4, 6..10
2011	338 (46.4%)	102 (14.0%)	0 (0%)	289 (39.6%)	0 (0%)	0 (0%)
2012	561 (76.9%)	19 (2.6%)	2 (0.3%)	74 (10.2%)	73 (10.0%)	0 (0%)

In Table I, a prediction example for the country Chad is given for the years 2011 and 2012. The permutation has been performed by usage of all six indicators and probability ranges split by tertiles. The prediction was based on indicators from 2002 to 2010. Together with reference MEPV-data for the same time span a SVM has been trained using the preferred method and kernel function as described in the next section. The reference MEPV-levels for the years 2002 to 2004 have been zero, for 2005 one and for 2006 to 2010 three. The table shows that the predicted indicators of Chad tend to be classified primarily with MEPV-level zero and secondarily with MEPV-level three for the years 2011 and 2012. The real MEPV-levels for these years were zero.

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-371-1

VI. KERNEL EVALUATION

As discussed before, an important aim is to find a well suited SVM kernel function to achieve a proper classification of indicator tuples in terms of MEPV-levels. For this, a separate optimization program has been written in C# instrumenting prior implemented prediction routines.

Five typical kernel functions have been selected for evaluation. Beside the two standard functions for linear and quadratic separation, the *Sigmoid*-, *Cauchy*- and *Hyperbolic Secant*-kernels have been chosen, based upon their principal properties, see [8][13]. These three non-standard kernel functions comprise scaling parameters, which have major influence on the classification results and thus need to be determined within a calibration process. The *Sigmoid*-kernel further represents the learning function of a two-layer-perceptron of a NN.

Because of the prior mentioned problem of over-classification, the kernels should not be evaluated by just using the training dataset as only reference to obtain the classification error. As an example, Table II shows evaluation results of the told kernels, trained with data from 2002 to 2010. The parameters for the three non-standard kernel functions have been optimized by minimizing the classification error of the training data. The table contains the classification errors and the standard deviations with respect to the reference MEPV-data for the years 2002 to 2010. The trained kernels have been used to classify the predicted indicator data for the years 2011 and 2012 using tertiles for the probability ranges of all six PC values. The gained results for both years, in form of the mean predicted MEPV-level for each country, are compared to the real MEPV-levels using their standard deviation. As it can be seen, with the *Cauchy*- and *Hyperbolic Secant*-Kernel it is possible to gain a 100 % fit for the training data, but the classification of the predicted data is worse than for the standard linear kernel. Using the quadratic kernel, the classifier cannot be trained at all.

TABLE II. SVM KERNEL PERFORMANCE
(OPTIMIZED FOR MINIMUM MISCLASSIFICATION OF TRAINING DATASET)

	Training (2002-2010)				
	linear	***quadratic***	***Sigmoid (0.05,-2.1)***	***Cauchy (0.0039)***	***Hyperbolic Secant (2.7)***
Class. Error	11.1 %	n/a[a]	10.0%	0%	0%
Std. Dev.	1.00	n/a[a]	0.98	0	0
	Classification of Prediction (2011-2012)				
Std. Dev.	0.85	n/a[a]	0.89	1.32	1.12

a. SVM training failed to find a solution

So, the goal must be to optimize the kernel function parameters by a minimization of the prediction error. The drawback of this kind of calibration is that the predicted data comprises additional prognosis uncertainties, influencing the accuracy of the found optimum. Due to a yet too small amount of available reference data, investigations on the required amount of years used for training and prediction with respect to the achieved overall prediction performance cannot be faithfully done so far.

In Table III, the results of using the method for minimizing the prediction errors are given. It can be seen that the *Cauchy*- and *Hyperbolic Secant*-kernels perform essentially better in the classification of the predicted values by conserving a small classification error of the training data. Especially the *Hyperbolic Secant*-kernel can be regarded as the kernel of choice for the current problem.

TABLE III. SVM KERNEL PERFORMANCE
(OPTIMIZED FOR MINIMUM MISCLASSIFICATION OF PREDICTED DATASET)

	Training (2002-2010)				
	linear	***quadratic***	***Sigmoid (0.05, -2.4)***	***Cauchy (0.71)***	***Hyperbolic Secant (0.56)***
Class. Error	11.1%	n/a[a]	10.1%	4.3%	1.3%
Std. Dev.	1.00	n/a[a]	0.96	0.59	0.39
	Classification of Prediction (2011-2012)				
Std. Dev.	0.85	n/a[a]	0.84	0.63	0.61

a. SVM training failed to find a solution

As of the currently quite small data basis, the kernel functions might perform slightly different when data from future years will be available. Further, there even might exist some more suitable, especially custom kernels. Regarding the potentially unlimited function space, an exhausting evaluation is not feasible anyway.

VII. SHORT TERM PROGNOSIS RESULTS

Applying the *Hyperbolic Secant*-Kernel with optimized kernel parameters, a one and a two year prognosis have been computed with respect to the last year of different time spans, which have been used for SVM training. The results are set in relation to the real values as mean and standard deviations through all countries and are listed in Table IV for each time span. Because data is currently available only until 2012, the last time span ends 2010 to remain reference data for up to two years.

TABLE IV. SUMMARIZED RESULTS FOR MEPV-LEVEL PROGNOSIS AS DEVIATIONS TO REAL DATA

	2002-2005	***2002-2006***	***2002-2007***	***2002-2008***	***2002-2009***	***2002-2010***
+ 1 year	0.19 ±0.93	0.01 ±0.88	0.05 ±0.56	-0.04 ±0.76	0.01 ±0.47	0.06 ±0.59
+ 2 years	-0.16 ±1.03	0.07 ±1.07	0.11 ±0.71	-0.05 ±0.92	0.07 ±0.84	0.03 ±0.62

It can be seen that the mean deviations are all close to zero, while the standard deviations range between 0.5 and 1.0 MEPV-levels. This can be regarded as acceptable by remarking that longer time periods also tend to imply smaller deviations and thus produce more truthful results. Due to too little data, yet undeterminable is at which point the length of the time span used for training can be regarded as sufficient. It can also be observed, that there is only a little increase in the deviations comparing the two with the one year prognosis.

Copyright (c) IARIA, 2014.

VIII. CONCLUSION

A method was presented to reconstruct probable societal political violence levels for countries which are not contained in the current MEPV-dataset using SVMs and world governance indicators. Further, the approach was extended to perform a prognosis of near future MEPV-levels. Therefore, a method has been described how to predict the underlying set of indicators by applying stochastic simulation and covering internal correlation effects using a PCA and a PC-transformation. While the forecast model itself is yet simple, as representing some kind of linear extrapolation, it provides an easy mechanism to obtain probabilities by stochastic modeling. As things usually do not develop just linear, a more sophisticated forecast model might provide more convenient results by e. g., also regarding interstate spillover effects which is not regarded so far. Finally, the current solution also delivers a reference model to test future models against.

An investigation of possible SVM kernels showed that the *Hyperbolic Secant*-kernel performed best by providing a classification error for data reconstruction and prediction with a standard deviation of about a half of an MEPV-level with respect to the real levels for the years 2011 and 2012.

As stated in the introduction, the targeted prediction time frame spans five years. Due to yet too few years for which data is available, only a time span of two years has been used for SVM kernel calibration and evaluation. In future, an extension to up to a five year time span is intended. For the same reason, an objective measure of the performance of the proposed method via a validation against past data is only limited possible.

ACKNOWLEDGMENT

Thanks are directed to the chair of *International Politics* from the *Department of Political Science, University of Federal Armed Forces* in Munich for providing guidance on underlying societal political aspects. Special thanks go to *Philipp Klüfers* who suggested the usage of the described datasets.

REFERENCES

[1] Federal Ministry of Defence (Germany), White Book, 2006.

[2] NATO, "The Alliance's Strategic Concept," Lisbon, 2010.

[3] D. Kaufmann, A. Kraay, and M. Mastruzzi, "The Worldwide Governance Indicators: Methodology and Analytical Issues," World Bank Policy Research Working Paper No. 5430, September 2010.

[4] J. Shlens, "A Tutorial on Principal Component Analysis," The Computing Research Repository, April 2014.

[5] M. Rychetsky, "Algorithms and Architectures for Machine Learning based on Regularized Neural Networks and Support Vector Approaches," Shaker Verlag GmbH Germany, December 2001.

[6] B. Schölkopf and A. J. Smola, "Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning)," The MIT Press, January 2002.

[7] S.S. Wilks, "Multivariate statistical outliers," Sankhya, Series A, 25, pp. 407–426, 1963.

[8] A. Chaudhuri, K. De, and D. Chatterjee, "A comparative study of kernels for the multi-class support vector machine," Fourth International Conference on Natural Computation, 2008, Vol. 2, pp. 3-7.

[9] O. J. de Groot, M. D. Rablen, and A. Shortland, "Gov-aargh-nance – "even criminals need law and order", " CEDI Discussion Paper Series 11-01, Centre for Economic Development and Institutions, Brunel University, April 2011.

[10] Center for Systemic Peace, "Major Episodes of Political Violence, 1946-2013," http://www.systemicpeace.org, retrieved: July 2014.

[11] World Bank, "Worldwide Governance Indicators, 1996-2012" http://www.worldbank.org, retrieved: July 2014.

[12] Martin Sewell, "Support Vector Machines (SVMs), " http://www.svms.org/, retrieved: July 2014.

[13] C. R. Souza, "Accord.net framework," 2013, http://accord-framework.net, retrieved: July 2014.

[14] The MathWorks Inc., "MATLAB R2014a", Natick, MA, 2000.

[15] S. Samarasinghe, "Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition", Auerbach Publications, 2006.

[16] G. McLachlan, "Discriminant Analysis and Statistical Pattern Recognition", John Wiley & Sons, 2004.