

EVIDENCE FOR THE CORRELATION BETWEEN CONFLICT RISK INDICATORS GCRI AND FSI USING DEEP LEARNING

Vera Kamp¹, JP Knust¹, Reinhard Moratz^{2,1}, Kevin Stehn¹ and Sören Stöhrmann¹

¹Data42 GmbH, Gotenstrasse 18, Hamburg, Germany

²Institute for Geoinformatics. University of Münster,
Heisenbergstraße 2, 48149 Münster, Germany

ABSTRACT

Data mining enables an innovative, largely automatic meta-analysis of the relationship between political and economic geography analyses of crisis regions. As an example, the two approaches Global Conflict Risk Index (GCRI) and Fragile States Index (FSI) can be related to each other. The GCRI is a quantitative conflict risk assessment based on open source data and a statistical regression method developed by the Joint Research Centre of the European Commission. The FSI is based on a conflict assessment framework developed by The Fund for Peace in Washington, DC. In contrast to the quantitative GCRI, the FSI is essentially focused on qualitative data. Both approaches therefore have closely related objectives, but very different methodologies and data sources. It is therefore hoped that the two complementary approaches can be combined to form an even more meaningful meta-analysis, or that contradictions can be discovered, or that a validation of the approaches can be obtained if there are similarities. We propose an approach to automatic meta-analysis that makes use of machine learning (data mining). Such a procedure represents a novel approach in the meta-analysis of conflict risk analysis.

KEYWORDS

Data Science, Deep Learning, Conflict Risk Prediction

1. INTRODUCTION

Data mining enables an innovative, largely automatic meta-analysis of the relationship between political and economic geography analyses of crisis regions. As an example, the two approaches Global Conflict Risk Index (GCRI) and Fragile States Index (FSI) can be related to each other. The GCRI is a quantitative conflict risk assessment based on open source data and a statistical regression method developed by the Joint Research Centre of the European Commission. The FSI is based on a conflict assessment framework developed by The Fund for Peace in Washington, DC. In contrast to the quantitative GCRI, the FSI is essentially focused on qualitative data. Both approaches therefore have closely related objectives, but very different methodologies and data sources. It is therefore hoped that the two complementary approaches can be combined to form an even more meaningful meta-analysis, or that contradictions can be discovered, or that a validation of the approaches can be obtained if there are similarities. We used machine learning (data mining) based on so-called multi-layered perceptrons. Perceptrons are the central and classical approach within the field of so-called artificial neural networks (ANN). These perceptrons are typically trained by backpropagation. Backpropagation belongs to the group of supervised learning methods and is applied to multi-layered perceptrons in that an external teaching function

knows the desired output, the target value, for a sample of inputs. This sample is also known as a training set. In the next section, we briefly summarize the properties of this robust and proven data mining approach.

2. THE STRUCTURE OF PERCEPTRONS

The single-layer perceptron can be viewed as a building block for Multi-layer Perceptrons. A single-layer perceptron maps an input vector an output vector with. The mapping behaviour of the perceptron is defined by its weight matrix (a matrix), its threshold vector and its threshold function. A multi-layer perceptron can be understood as a series of several single-layer perceptrons. There is a net representation of this calculation process. The vectors correspond to one layer of processing nodes (in the net view the input vector is also a node layer). These processing nodes are also called neurons.

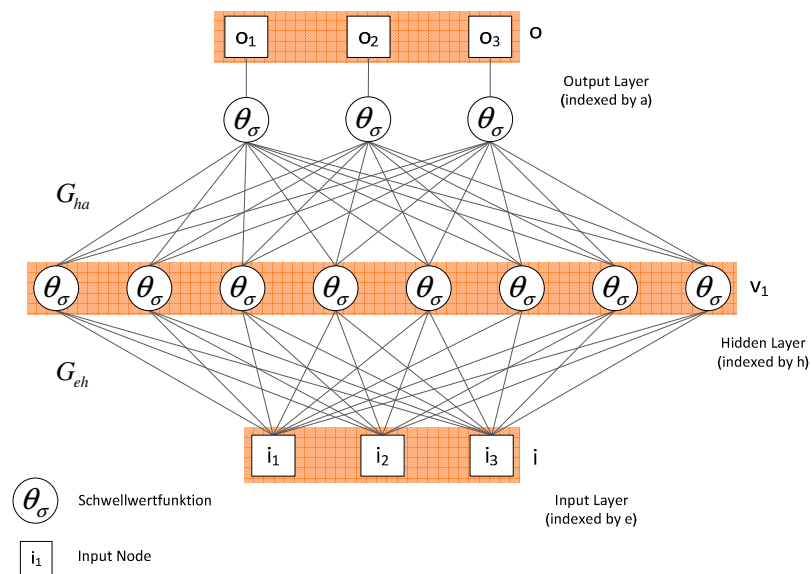


Figure 1: Multilayer perceptron with a single hidden layer

2.1. BACKPROPAGATION TRAINING ALGORITHM

The backpropagation method is used to train multi-layer perceptrons with a sigmoid threshold function. For a given training set, the total error on this sample should be minimal. Since the backpropagation procedure uses the steepest descent method, a continuous expression must be found that has a minimum at a potential solution of the learning problem. Therefore a quadratic error function is used. This error function only has a minimum value of 0 if there is an error-free solution to the learning problem. The sequence of the backpropagation procedure follows this scheme. At the start all coefficients of weight matrices and threshold vectors get an assignment with pseudo random numbers with values between -0.5 and 0.5. Then each learning cycle involves an iteration that ensures that after each step the overall error value becomes smaller. Then there is a termination condition which detects if there is a local minima found which potentially could be a good solution to the learning problem related to the training set. Numerically this final state, at which the gradients go to 0 and no modification of the weights takes place, corresponds to a minimum of the error function. Obviously, in an unfavorable case, this minimum can be a poor local minimum (see Tesi92[9]). In order to accelerate the learning process, a so-called momentum term is usually inserted, which includes weight changes from earlier steps (see Rumelhart86[21], Chapter 8).

3. DEEP LEARNING FOR CROSS VALIDATION OF GCRI AND FSI

From the perspective of machine learning the training of a model that has complex inner representations is a challenge. That is why training a multi-layer perceptron with multiple hidden layers is called deep learning. This refers to inner representation that are automatically build during the learning process. In our case we only have a single hidden layer. In further experiments we will use multiple hidden layers but the training set for our first experiment was limited so using a single hidden is adequate and typically one would also refer to this method as deep learning even if the full potential of deep learning would be only used in later experiments which are sketched below.

In our first experiment we created an objective function based on the numerical indicators of the Global Conflict Risk Index (GCRI) as inputs. These numerical indicators are already standardized to the interval 0.0 to 10.0. Since multi-layer perceptrons with their activity levels can assume values between 0.0 and 1.0, the GCRI values must be multiplied by 0.1 to obtain valid input values for a perceptron. A series of input values from the GCRI then correspond to an input vector (see section 2) of the perceptron. This input vector is given a series of target values, the target vector, as the desired output. In the first experiment we took the data from the Fragile-State-Index/Fund for peace (source: <http://fsi.fundforpeace.org/data>) as target values. In this first experiment, a target vector of the same country and year was assigned to the input vector in relation to the year of the country. The names of the countries from GCRI and Fragile State Index must be assigned to each other accordingly. In our experiment, we used the country coding of the ISO-3166-1 coding list. The target vectors are then also normalized to the interval from 0.0 to 1.0. For training the data of all countries of the years 2006 to 2016 were used. The prediction quality was then determined with the data of all countries of the years 2017 and 2018. This data is also known as the validation sample.

We started with a multi-layer perceptron with only one hidden layer. We started with five hidden nodes and then went with five steps to a previously calculated maximum number of hidden nodes.

The parameter was calculated according to the following rule of thumb: One calculates two bits of information storage capacity per edge weight. A network with too many hidden nodes has so much storage capacity that it can memorize the data content of the training data set and then in the worst case generalize it hardly rationally to new/unknown data. So the choice of the maximum number of hidden nodes can be automatized.

After the multi-layer perceptron backpropated thousands of input-target pairs from the training set, the mean error on the original scale of fragile state indices with the interval from 0 to 10 was 0.71 for the validation sample. The corresponding net then has 25 hidden nodes. A prediction with a mean error of only 0.77 is already achieved with only 5 hidden nodes. This very impressive result and its application aspects are discussed in the next section.

As a technical detail you can see on the diagram on the next page that with a very large number of hidden nodes the effect described above with the capacity calculation actually occurs. A very high storage capacity of the network can lead to an over-adaptation to the training sample only and the prediction quality measured at the validation sample becomes significantly worse.

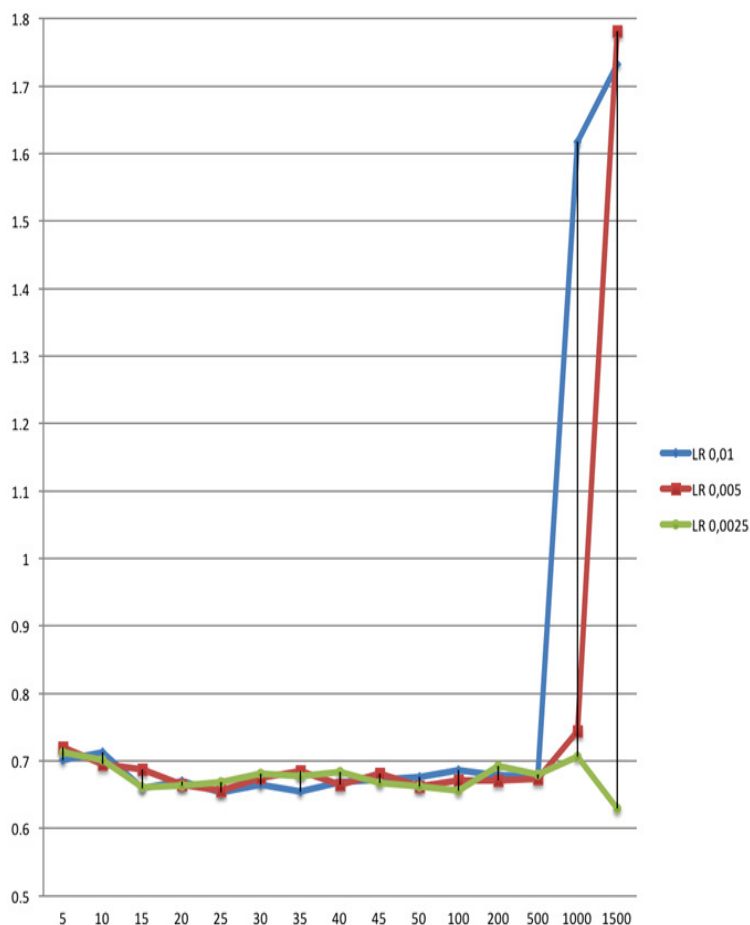


Figure 2: A very high storage capacity of the network can lead to an over-adaptation to the training sample only and the prediction quality measured at the validation sample becomes significantly worse.

4. DISCUSSION AND CONCLUSIONS

The relatively small error in predicting from the Global Conflict Risk Index 2018 data to the Fragile State Index 2018 data means that not only can the perceptron reproduce the given data from 2006 to 2016, but it has learned regular relationships between GCRI and FSI that can also be successfully applied to data not included in the training sample, such as the 2018 data. For example, for the year 1999, GCRI data can be used to generate hypothetical FSI data. As the Fragile State Index only dates back to 2005, this is already an interesting added value.

With our data mining approach, we have succeeded in correlating the two approaches Global Conflict Risk Index (GCRI) and Fragile States Index (FSI). We were able to predict the FSI Index relatively accurately with a mean deviation of 0.7 from the GCRI Index for the same year. This also clearly shows that the quantitative approach using open source data from the GCRI and the qualitative conflict assessment framework of the FSI, which is based on expert ratings, correlate strongly. This finding strengthens confidence in the meaningfulness of both approaches. Both approaches thus seem to generate comparable descriptions through their closely related objectives, although they are based on very different methodologies and data sources. Both approaches thus support each other, as we were able to show with our approach to automatic meta-analysis using machine learning (data mining). Our method thus represents a novel approach in the meta-analysis of conflict risk analyses.

Furthermore, there are interesting possibilities to extend our approach. For example, you can train predictions from the present into the near future. For example, use GCRI indicators for 2017, 2018, 2019 of a country to predict FSI indicators for 2020 of that country (i.e. which do not yet exist). For this purpose, Global Conflict Risk Index data of the years can be used as input vectors and Fragile State Index of the year as target vectors. Our innovative approach can therefore be extended directly to predict future trends.

REFERENCES

- [1] E. B. Baum: On the Capabilities of Multilayer Perceptrons, *Journal of Complexity*, Vol.4, No.3, (1988).
- [2] E. B. Baum, D. Haussler: What Size Net gives Valid Generalisation ? *Advances in Neural Information Processing Systems*, D. Touretzky, Ed., Morgan Kaufmann, San Mateo, CA, Seite 81-90 (1989).
- [3] E. B. Baum: On Learning a Union of Half Spaces, *Journal of Complexity*, Band 6, Seite 67-101 (1990).
- [4] E. B. Baum: When Are K-Nearest Neighbour and Backpropagation Accurate for Feasible Sized Sets of Examples, in [?]
- [5] E. B. Baum: A Polynomial Time Algorithm That Learns Two Hidden Unit Nets, *Neural Computation* 2, 510-522, (1991).
- [6] A. Blum, R. Rivest: Training a 3-Node Neural Network is NP-complete, in COLT88[?] Seite 211-218 (1988).
- [7] T. M. Cover: Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, *IEEE Trans. Electron. Comput.* EC-14, 326-334.
- [8] De Groeve T, Hachemer P, Vernaccini L. The global conflict risk index (GCRI): a quantitative model. *Concept and Methodology*. 2014.
- [9] M. Gori, A. Tesi On the Problem of Local Minima in Backpropagation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Band 14, Nr. 1, (1992).
- [10] Haken, Nate, J. J. Messner, Krista Hendry, Patricia Taft, Kendall Lawrence, Laura Brisard, and Felipe Uman?a. 2014. *Failed State Index 2014: The Book*. Washington, D.C.: Fund for Peace.
- [11] D. O. Hebb: *The Organisation of Behaviour*, New York: Wiley (1949).
- [12] J. Hertz, A. Krogh, R. G. Palmer Introduction to the Theory of Neural Computation, Addison-Wesley, Redwood City, Ca, (1991).
- [13] G. E. Hinton: Wie neuronale Netze aus Erfahrung lernen, in *Spektrum der Wissenschaft*, November-Ausgabe, Heidelberg, (1992).
- [14] J. S. Judd: *Neural Network Design and the Complexity of Learning*, MIT Press, Cambridge, MA (1990).
- [15] J. L. McClelland, D. E. Rumelhart *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, A Handbook of Models, Programs, and Exercises*, Cambridge, MA: MIT Press 1988.
- [16] W. S. McCulloch, W. Pitts: A Logical Calculus of the Ideas Imanent in Nervous Activity, *Bull. Math. Biophys.* 5, 115-133 (1943).

- [17] M. Minsky, S. Papert: *Perceptrons*, Cambridge, MA: MIT Press 1969.
- [18] G. J. Mitchison, R. M. Durbin: *Bounds on the Learning Capacity of some Multi-Layer Networks*, *Biol. Cybern.* Vol. 60, No.5, (1989).
- [19] B. K. Natarajan: *Machine Learning, A Theoretical Approach*, Morgan Kaufmann, San Mateo, CA, (1991).
- [20] F. Rosenblatt: *Principles of Neurodynamics*, New York: Spartan (1962).
- [21] D. E. Rumelhart, J. A. McClelland: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Cambridge, MA: MIT Press 1986.
- [22] L. G. Valiant: *A Theory of the Learnable*, *Communications of the ACM*, Band 27, Nr.11. Seite 1134-1142
- [23] V. Vapnik: *Estimation of Dependences Based on Empirical Data*, Springer-Verlag, New York (1982).
- [24] G. Widrow, M. E. Hoff: *Adaptive Switching Circuits*, *Institute of Radio Engineers, convention Record, Part 4*, 96-104 (1960).