International Journal of Artificial Intelligence & Applications (IJAIA) Vol.10, No.4, July 2019

Data Mining for Integration and Verification of Socio-Geographical Trend Statements in the Context of Conflict Risk

Vera Kamp, Jean-Pierre Knust, Reinhard Moratz, Kevin Stehn, Soeren Stoehrmann

{vera.kamp, jp, kevin.stehn, soeren.stoehrmann}@data42.de
data42 GmbH
moratz@uni-muenster.de
University of Muenster

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 from systematic interviews with experts.
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 analyses.

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 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 analyses.

DOI : 10.5121/ijaia.2019.10401

In our context, the use of so-called multi-layered perceptrons is a suitable approach. 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 multilayered 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 mathematical properties of this robust and proven data mining approach.

2 Principles of Multi-Layer Perceptrons

A single-layer perceptron realizes a mapping from an input vector $\mathbf{i}$ with e components to an output vector $\mathbf{o}$ with a components. The mapping function of the perceptron is determined by its weight matrix $\mathbf{G}$ (a $e \times a$ matrix), its threshold vector $\mathbf{s}$ (a vector with a components) and its threshold function θ The techniques discussed in this work use the threshold function θ_σ. The threshold function θ_σ is thus defined:

$$\theta_\sigma \left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ j \\ \vdots \end{bmatrix} \right) = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_j \\ \vdots \end{bmatrix}, \qquad y_j = \frac{1}{1 + e^{-x_j}}$$

Then $\mathbf{o}$ is calculated from $\mathbf{i}$ as follows:

$$\mathbf{o} = \theta(\mathbf{G}\mathbf{i} + \mathbf{s})$$

A multi-layered perceptron can be constructed as a sequence of several coupled single-layer perceptrons. Let $\mathbf{i}$ be the input of the first single-layer perceptron, its output is as $\mathbf{v}_1$ (a vector with h_1 components). The input of the second perceptron is then $\mathbf{v}_1$, whose output is $\mathbf{v}_2$, and so on. The output of the last single-layer perceptron in this chain is is $\mathbf{o}$ and thus also the output of the entire multilayer perceptrons. The number of layers is z. The weight matrices of the perceptrons are $\mathbf{G_1}, \mathbf{G_2}, \ldots, \mathbf{G_z}$. their threshold vectors $\mathbf{s_1}, \mathbf{s_2}, \ldots, \mathbf{s_z}$. The output vector l ”a ”is then calculated like this:

$$\begin{aligned} \mathbf{v}_1 &= \theta\left(\mathbf{G_1}\mathbf{i} + \mathbf{s_1}\right) \\ \mathbf{v}_2 &\quad \theta\left(\mathbf{G_2}\mathbf{v}_1 + \mathbf{s_1}\right) \\ &\vdots \\ \mathbf{o} &= \theta\left(\mathbf{G_z}\mathbf{v}_{z-1} + \mathbf{s_z}\right) \end{aligned}$$

$$\mathbf{o} = \theta(\mathbf{G_z} \ldots \theta\left(\mathbf{G_2}\,\theta\left(\mathbf{G_1}\mathbf{i} + \mathbf{s_1}\right) + \mathbf{s_2}\right) \ldots + \mathbf{s_z})$$

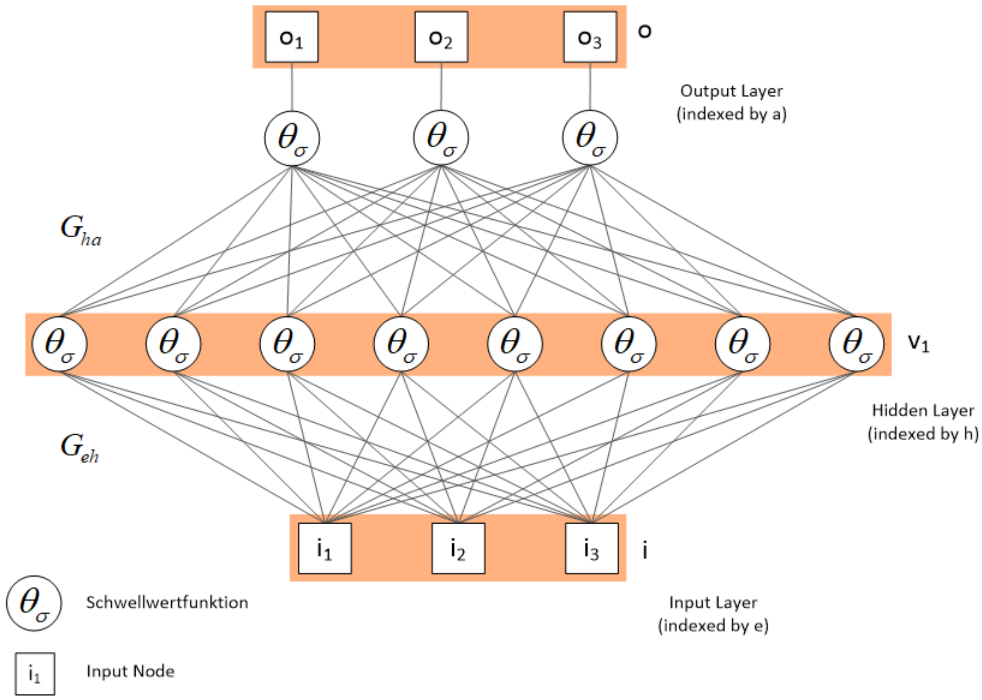

Fig. 1. A multi-layered perceptron and its parameters.

There is a network representation of this calculation process. Thereby vectors $\mathbf{i}, \mathbf{v}_1, \ldots, \mathbf{v}_{z-1}, \mathbf{o}$ are each viewed as layers of processing nodes (in the network perspective, the input vector is also a node layer). Then each component of these vectors is viewed as processing node. These processing nodes are also called neurons.

2.1 Backpropagation

With the backpropagation training procedure, multi-layer perceptrons with the threshold function θ_σ can be adapted to a training set (e.g. supervised machine learning). The procedure is carried out on a two-stage network with a The feed node is auto-explained. This network has the parameters $\mathbf{G}_1$,$\mathbf{G}_2$,$\mathbf{s}_1$ and $\mathbf{s}_2$. Since the threshold function θ_σ is used, the value is of the output node is a real number from the interval from 0 to 1. For a given Me series (y^m, x^m), the total error is to be set to of this sample $f_\mathrm{k}((y^m, x^m))$ must be minimal. Since the backpropagation procedure is the method of the steepest descent is used, a constant expression must be found, which is has a minimum at $f_\mathrm{k}((y^m, x^m)) = 0$.

Use a quadratic Error function J_{BP}.

$$\mathrm{J}_{BP}\left(\mathbf{G}_1, \mathbf{G}_2, \mathbf{s}_1, \mathbf{s}_2\right) = \sum_{i=1}^{m} \frac{1}{2} \cdot \left(y_i - \left(\theta_\sigma\left(\mathbf{G_2}\, \theta_\sigma\left(\mathbf{G_1}\mathbf{x}_i + \mathbf{s_1}\right) + \mathbf{s_2}\right)\right)\right)^2$$

This error function has a minimum value of 0, if there is an error-free solution of the loading problem. The backpropagation procedure follows this schema:

Startup:

All components from $\mathbf{G}_1, \mathbf{G}_2, \mathbf{s}_1, \mathbf{s}_2$ get pseudo-random numbers with values between -0.5 and 0.5.

Iteration

$$\mathbf{G}_1^{New} := \mathbf{G}_1^{old} + \lambda \cdot \frac{\partial \mathrm{J}_{BP}\left(\mathbf{G}_1^{alt}, \mathbf{G}_2^{alt}, \mathbf{s}_1^{alt}, \mathbf{s}_2^{alt}\right)}{\partial \mathbf{G}_1^{alt}}$$

$$\mathbf{G}_2^{new} := \mathbf{G}_2^{old} + \lambda \cdot \frac{\partial \mathrm{J}_{BP}\left(\mathbf{G}_1^{alt}, \mathbf{G}_2^{alt}, \mathbf{s}_1^{alt}, \mathbf{s}_2^{alt}\right)}{\partial \mathbf{G}_2^{alt}}$$

$$\mathbf{s}_1^{new} := \mathbf{s}_1^{old} + \lambda \cdot \frac{\partial \mathrm{J}_{BP}\left(\mathbf{G}_1^{alt}, \mathbf{G}_2^{alt}, \mathbf{s}_1^{alt}, \mathbf{s}_2^{alt}\right)}{\partial \mathbf{s}_1^{alt}}$$

$$\mathbf{s}_2^{new} := \mathbf{s}_2^{old} + \lambda \cdot \frac{\partial \mathrm{J}_{BP}\left(\mathbf{G}_1^{alt}, \mathbf{G}_2^{alt}, \mathbf{s}_1^{alt}, \mathbf{s}_2^{alt}\right)}{\partial \mathbf{s}_2^{alt}}$$

This ensures that after each step under the condition $\lambda \rightarrow 0$ the inequality $\mathrm{J}_{BP}\left(\mathbf{G}_1^{new}, \mathbf{G}_2^{new}, \mathbf{s}_1^{new}, \mathbf{s}_2^{new}\right) \leq \mathrm{J}_{BP}\left(\mathbf{G}_1^{alt}, \mathbf{G}_2^{alt}, \mathbf{s}_1^{alt}, \mathbf{s}_2^{alt}\right)$ applies.

Termination condition:

$$\mathrm{J}_{BP}\left(\mathbf{G}_1^{alt}, \mathbf{G}_2^{alt}, \mathbf{s}_1^{alt}, \mathbf{s}_2^{alt}\right) - \mathrm{J}_{BP}\left(\mathbf{G}_1^{new}, \mathbf{G}_2^{new}, \mathbf{s}_1^{new}, \mathbf{s}_2^{new}\right) < \text{Tolerance value}$$

The final state, where the gradients amountm ”a ”sig go towards 0 and no modification of the weights takes place any more corresponds to a minimum of the error function. This minimum can be set in a In our case, it must be a local minimum (see Tesi92[10]).

In order to speed up the loading process we usually use a the so-called momentum term, which allows for weight ”changes earlier steps included (see Rumelhart86[23], Chapter 8).

3 Numerical data mining for cross-validation of GCRI and FSI

In a first experiment, we created an objective function based on the numerical indicators of the Global Conflict Risk Index (GCRI) as input. These numerical indicators are already standardized to the interval 0.0 to 10.0. Since multi-layer perceptrons with their activation levels can assume values between 0.0 and 1.0, the GCRI values must be multiplied by 0.1 to obtain regular input values for a perceptron. A series of input values from the GCRI then corresponds to an input vector **o** (see section 2) of the perceptron. This input vector is given a series of target values, the target vector as a desired output according to the corresponding Fragile-State-Index component. In the first experiment we took the data from the Fragile-State- Index/Fundforpeace as target values (source: http://fsi.fundforpeace.org/data). In this first experiment, the input vector is assigned with respect to year y_i of the country l_j a target vector of the same country l_j and of the same year y_i. For this purpose the names of the countries must be matched from GCRI and Fragile-State-Index. In our experiment we used the ISO country coding. The target vectors are then also normalized to the interval from 0.0 to 1.0.

We started with a multi-layer perceptron with only one hidden layer. We started with five hidden nodes and then went with increments of 5 to a previously calculated maximum number of h hidden nodes. The parameter h was calculated according to the following rule of thumb: One calculates per edge weight two bits information memory capacity. The dimension of the input vector is e , the dimension of the output vector and at the same time of the target vector is a (see section 2).

Thus the net has calculated approximatively $i * h + h * o$ edge weights and then a memory dependence $C_{net} = (2 * i * h + h * o)$ Bit. The training set has approximatively calculated a data content of $D_{train} = number\_of\_the\_example\_vector\_pairs * * o * 4 * o$ bit. A network with too many hidden nodes has so much data storage capacity C_{net}, that it contains the data content D_{train} of the Training data set "can learn by heart" can and then in the worst case hardly generalize to new/unknown data. So h is chosen so that $C_{net} < D_{train}$ applies.

After the multi-layer perceptron backpropated thousands of such value pairs, 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.

As a first practical benefit we can deduce for the year 1999 from GCRI data generate hypothetical fragile state index estimates. As the fragile state index is only available until 2005, this is an added value.

In the next steps, predictions from the near present into the near future are aimed at. For example, the Global Conflict Risk Index can use numerical indicators for 2014, 2015, 2016 of a country l_j, to predict the fragile state index for 2017 of this country l_j (i.e. it doesn't even exist yet). GCRI data of the years y_i, y_{i+1}, y_{i+2} are used as input vector and fragile state index. of the year y_{i+3} as target vectors. The multi-layered perceptron then has the triple the entrance width.

In the third experiment, we will specify the difference of two consecutive years in the fragile state index as the target value. Then an increase or decrease of the state of a country with respect to instability ”at is trained as a prediction. In the fourth experiment, pairs of input and target vectors are identified in the training set that indicate a deterioration of a country. These risk pairs are added in three copies to the original training set. Thus, the weight of risk predictions is weighted fourfold. This corresponds to a one-sided error philosophy. Ideally, the system then becomes good at predicting possible risk situations and is particularly sensitive to negative developments. The system can also generate ”false alarms”, but these are considered to be much less dangerous than if the system were to Risk situations ”overlooks”.

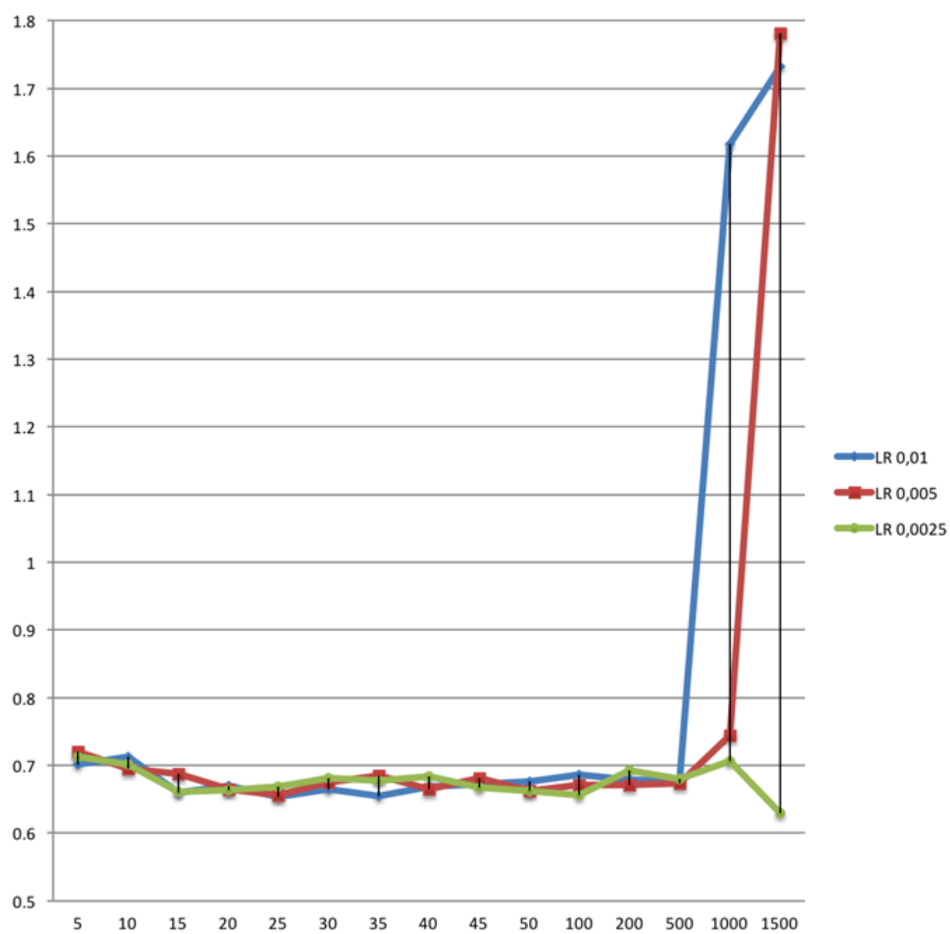

Fig. 2. Overfitting reduces the generalization performance.

4 Discussion and outlook

The relatively small error in predicting from the Global Conflict Risk Index 2014 data to the Fragile State Index 2014 data means that not only can the perceptron reproduce the given data from 2006 to 2012, 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 2014 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, page 81-90 (1989).
3. E. B. Baum: *On Learning a Union of Half Spaces*, Journal of Complexity, Volume 6, Page 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[**?**] Page 211-218 (1988).
7. T. M. Cover: *Geometrical and statistical properties of systems of linear inequalities with applications in pattern recogniton*, 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. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep learning.* MIT press, 2016.

10. M. Gori, A. Tesi *On the Problem of Local Minima in Backpropagation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 14, No. 1, (1992).
11. 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.
12. D. O. Hebb: *The Organisation of Behaviour*, New York: Wiley (1949).
13. J. Hertz, A. Krogh, R. G. Palmer *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, Ca, (1991).
14. G. E. Hinton: *Wie neuronale Netze aus Erfahrung lernen*, in Spektrum der Wissenschaft, November-Ausgabe, Heidelberg, (1992).
15. Marsland, Stephen. *Machine learning: an algorithmic perspective.* Chapman and Hall CRC, 2011.
16. J. S. Judd: Neural Network Design and the Complexity of Learning, MIT Press, Cambridge, MA (1990).
17. 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.
18. W. S. McCulloch, W. Pitts: *A Logical Calculus of the Ideas Imanent in Nervous Activity*, Bull. Math. Biophys. 5, 115-133 (1943).
19. M. Minsky, S. Papert: *Perceptrons*, Cambridge, MA: MIT Press 1969.
20. G. J. Mitchison, R. M. Durbin: *Bounds on the Learning Capacity of some Multi-Layer Networks*, Biol. Cybern. Vol. 60, No.5, (1989).
21. B. K. Natarajan: *Machine Learning, A Theoretical Approach*, Morgan Kaufmann, San Mateo, CA, (1991).
22. F. Rosenblatt: *Principles of Neurodynamics*, New York: Spartan (1962).
23. D. E. Rumelhart, J. A. McClelland: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Cambridge, MA: MIT Press 1986.
24. L. G. Valiant: *A Theory of the Learnable*, Communications of the ACM, Band 27, Nr.11. Seite 1134-1142
25. V. Vapnik: *Estimation of Dependences Based on Empirical Data*, Springer-Verlag, New York (1982).
26. G. Widrow, M. E. Hoff: *Adaptive Switching Circuits*, Institute of Radio Engineers, convention Record, Part 4, 96-104 (1960).