Data Normalization with Self-Organizing Feature Maps

Alfred Ultsch, Günter Halmans
Department of Computer Science, University of Dortmund
P.O. Box 500 000
D-4600 Dortmund 50, Germany
email: ultsch@exunido.irb.informatik.uni-dortmund.de

Abstract

The processing of empirical data brings forth the need to transform observed distributions to a normal distribution before they can be processed or analyzed furthermore. The selection of a fitting transformation is typically an expert task including trial and error. This paper presents a method to find a suitable transformation using a self-organizing feature map. The feature map's learning algorithm was suitably modified in order to predict the parameter for a transformation. First results show that the map is able to recall the training set almost exactly and, furthermore, that the model is able to generalize. Experiments with unknown distributions exhibit promising error rates.

1. Introduction

The processing of empirical data rises often the question how to transform this in order to make it comparable. In many cases, empirical data is not distributed, but has to be transformed to some sort of normal distribution by an apt transformation. In statistics, several transformations are known to achieve normality or at least a symmetric distribution [Hartung 87].The selection of an apt transformation for a given empirical distribution, however, is not trivial. It is often considered a trial and error process which takes an expert to perform appropriately [Schlittgen 90]. Often only the experience from many distributions with their corresponding structures and the required transformations make a qualified selection feasible.

This paper presents a method to find a suiting transformation by the usage of a self-organizing feature map. The reader of this paper is assumed to be in principle familiar with Kohonen's self-organizing feature maps [Kohonen 84]. Only the differences to the standard model are explained below (see chapter 3).

2. Data Transformation in Explorative Data Analysis

In order to classify data sets and their distribution, it is necessary to define some parameters which describe their characteristics. Empirical and theoretical distributions are, for example, characterized by median, mean, variance, percentiles or skewness. The values of the median and the mean of a standard distribution, for example, are equal because of the symmetric distribution form. The symmetric structure also involves that the distribution has no skewness [Hartung 87].

exponent p	... 3 ...	2 ...	1 ...	0,5 ...	ln ...	-0,5 ...	-1 ...	-2 ...
transformation	... x^3...	x^2...	x ...	$\sqrt{x}$ ...	ln(x) ...	$1/\sqrt{x}$ ...	1/x ...	$1/x^2$...

distribution form: left skewed <------> symmetric <-----> exponential <--------------------> right skewed

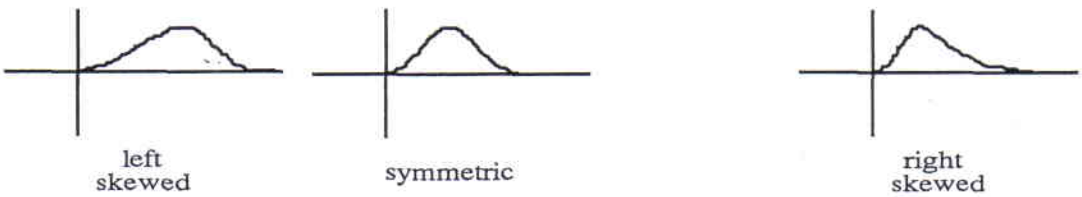


Figure 1: The ladder of power

A typical normalization of a data set often has the form x^p. The so-called "ladder of power" characterizes the quality of the different transformations [Hartung 87]. Figure 1 shows, which transformations are to be used in relation to the different structures of the distributions. A distribution skewed to the right needs a transformation exponent p less than one. A distribution skewed to the left needs a transformation exponent p greater than one. The ln-transformation can be inserted in the ladder of power in place of p = 0. For negative exponents p the order of the data is inverse. In this case the transformation $-(x+c)^p$ is used. The problem of transforming negative values can be avoided by adding a constant.

Given a distribution, the ladder of power tells whether the transformation exponent is greater or less than one. The selection of the exact value of the exponent is typically done by performing a number of transformations using different values for p and comparing the result with a normal distribution using, for example, a Q/Q-plot [Hartung 87]. The selection of a qualified exponent is typically a lengthy trail and error process [Schlittgen 90].

3. The network used

By using a connectionist model, the self-organizing feature map [Kohonen 84], we attempted to find an automatic process for the selection of the right exponent. Connectionist models like the self-organizing feature map claim that they can generalize and handle with noisy and incomplete data [Rumelhardt/McClelland 86]. They are able to make a plausible statement about missing components [Ultsch (Ed) 90][Ultsch et al 91]. The method described in this paper uses especially the last property. We used a Kohonen self organizing feature map with 32 x 32 units arranged in a rectangular grid. The input vectors were 10 dimensional: nine percentiles of a given distribution and the transformation exponent (see Figure 2).

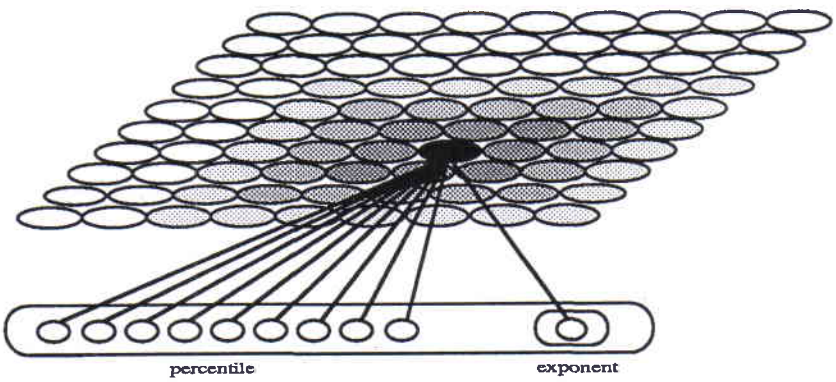


Figure 2: The modified Kohonen model

The learning phase differs from the traditional Kohonen learning phase. For calculating the unit distances only the first 9 components were considered. So the unit with the nearest weight vector in the sense of the Euclidean Distance is determined with respect to the percentiles exclusively. Therefore, the feature map arranges the different distributions according to the characteristics of the distribution. In the adaption phase, all components, including the exponent, are used to adapt the weight vector of a unit in the neighborhood of a unit. This modification allows the network to interpolate and predict the exponents for different distributions.

Given an empirical distribution, the percentiles are used to find the most similar distribution in the trained feature map. The net matches the input vector into the feature map and calculates the unit with the nearest weight vector. For this unit the 10th component, i.e. the exponent, is returned by the network as a plausible prediction of the exponent for normalizing the input distribution. The capability of generalization allows the net to classify a distribution which the training set did not contain.

4. The training data

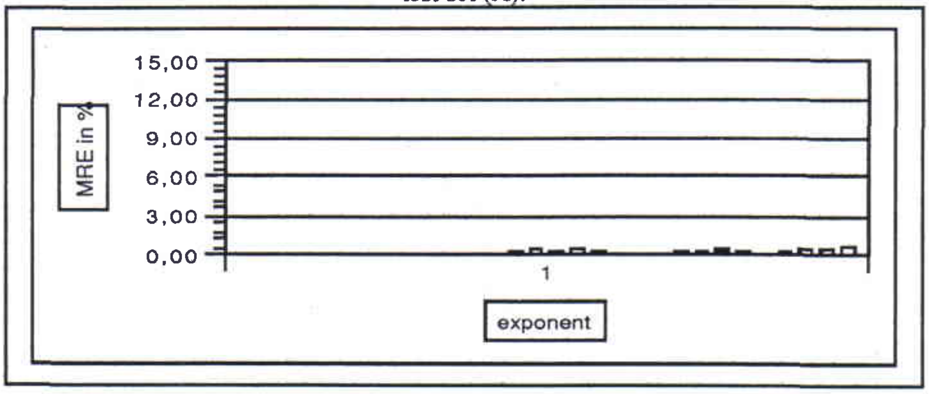
In order to get distributions with known transformation exponents we generated N(0,1) standard normal distributed data sets of 100 points each.These data sets were transformed using different exponents p. In this way we generated 30 differnent data sets with known transformation exponent. The exponents ranged from .1 to 9 and were exponentially distributed in this range.

The distributions were characterized through 10 components. The first nine components contained the 9 percentiles of the distribution, the 10th the exponent. The percentiles were given in percentage of the data range.

5. Results

The Kohonen network was implemented on a transputer system [Ultsch/Siemon 89]. With this implementation it is possible to experiment in a very efficient way. Additionally the training process can be animated and vizualized. The self-organizing feature map has been learned by presenting the training distributions about 250.000 times.

test set (A):



test set (B):

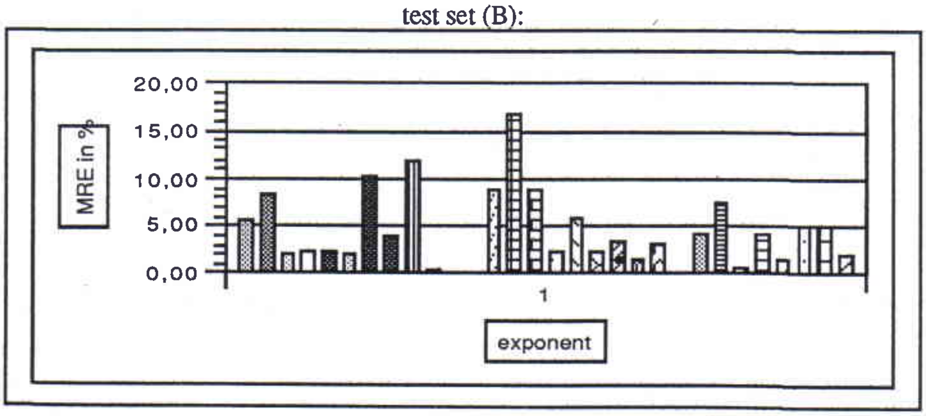


Figure 3: First results

To test our method, we used two different test sets. The first set (A) is equal to the training set. By checking the accuracy of the prediction of the training distributions we could see whether the net had been trained sufficiently. The second test set (B) consists of 30 new distributions with a known transformation exponent . This set allows to judge the accuracy of the generalisation. Figure 3 shows the mean relative error in percent for the three data sets.

The first experiment points out, that the net has learned very well. The mean relative error (MRE) in percent for all 30 distributions is 0.14 %. The results of the second test show that the mean relative error is typically below 4.5% with a standard deviation below 4.0 %. So the net is able to generalize to new distributions.

6. Discussion

A statistical method to estimate the transformation exponent uses a so called p-quantile coefficient[Schlittgen 90]:

$$g_p(m) = \frac{(x_{p-1}^m - \tilde{x}^m) - (\tilde{x}^m - x_p^m)}{x_{p-1}^m - x_p^m} \qquad \text{m: transformation} \qquad \text{p : quantile}$$

In the case of a symmetric distribution, the value of the coefficient is 0. If the coefficient is positiv, the distribution is skewed to the right.A negative coefficient implies that the distribution is skewed to the left.

Given a distribution, a transformation from the ladder of power is selected and applied. For the resulting distribution, the p-quantile coefficient is calculated. This process is repeated until the p-quantile coefficient is approximately zero. If the p-quantile coefficient is approching zero, it may, however, be that the resulting transformation is not correct because the data is exponentially distributed. This can only be checke using a different coefficient [Schlittgen 90].

In summary, the traditional method to find a suitable tranformation involves a lot of trial and error. It may be a time-consuming process involving a statistical expert. Our method, in contrast, draws heavily on the properties of a Kohonen feature map. New distributions are not only compared to the learned distributions but also to generalizations of the learned distributions. A most similar distribution is found and the generalized transformation exponent returned. Our preliminary results exhibit a promising low error rate for an automatic process of transformation parameter estimation.

7. Conclusion

In this paper we showed the principle possiblity to substitute the trial and error process for finding data transformation by using a self-organizing feature map. We generated different distributions with different skewness and trained a modified Kohonen self-organized feature map with a description of the data. First results point out, that the net is able to recall the training set almost exactly. Furthermore, the model is able to generalize to different transformations and to estmate the transformation parameter for unknown distributions with promising error rates.

References

[Hartung 87] Hartung, J.: *Statistics*,Oldenbourg, Munich 1987, in German

[Kohonen 84] Kohonen, T.:*Self-Organisation and Associative Memory*, Springer Verlag, Berlin 1984

[Fanihagh et al 90] Fanihagh, F., Lütgendorf, A. Mempel, M., Rossbach, P.,Schneider, B., Wegmann, F. *Knowledge acquisition for knowledge-based systems with connectionist models*, in [Ultsch(Ed.) 90], in German.

[Rumelhardt/McClelland 86] Rumelhardt, D.E.; McClelland J.L.: *Parallel Distributed Processing: Exploration in the Microstructure of Cognition, Volume 1: Foundations*, MIT Press, Cambridge (Massachusetts) 1986

[Schlittgen 90] Schlittgen, R.: *Introduction to Statistics*, Oldenbourg, Munich 1990, in German

[Ultsch (Ed) 90] Ultsch, A. (Ed.): *Coupling declarative and connectionist representations for Knowledge*, Reasearch Report No. 352, Department of Computer Science, University of Dortmund, 1990, in German

[Ultsch et al. 91] Ultsch, A., Halmans, G., Mantyk, R.: *CONKAT: A Connectionist Knowledge Acquisition Tool*, in: Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences, IEEE Computer Society Press, Los Alamitos, California 1991

[Ultsch/Siemon 89] Ultsch, A., Siemon, H.P.: *Exploraty Data Analysis: Using Kohonen Networks on Transputers* Department of Computer Science, Technical Report No. 932, University of Dortmund 1989