An Implied Rating Software System

Ventsislav Nikolov

Abstract. *The paper presents a mathematical approach to create credit rating scale and to classify financial institutions by it which is implemented as a model in a software system. The presented model is based on competitive trained neural network working in model building and classification stages. Thus, an individual point of view can be provided for any institution according to their available data.*

Keywords: *Implied rating, Classification, Self-organizing map*

1 INTRODUCTION

The credit ratings evaluating the credit worthiness of different obligators [2] are important data for the business and governments. When there is a financial and economic crisis the importance of the ratings produced by the rating agencies even raises because they influence the investors' decisions and corporate operations toward a given direction. The ratings determine the interest rate for the borrower which leads to different prices of loaning money. Here a method for automatic building of a credit ratings scale based on specific corporate and government data as well as determination of the credit rating of a given borrower is described. The method is implemented as a software module which can be integrated in a variety of software products.
The need of new methods for credit rating determination emerges because of the following reasons.

- The reaction of the rating agencies often is too slow and the market is dynamic. Sometimes even a default occurs of a company or institution while their rating is still classified as high one.
- The ratings are determined in long time intervals. Ratings available on daily basis could be a significant advantage.
- Sometimes the rating agencies are criticized of conflict of interests. They analyze the political environment, regulations, the ability to return already borrowed loans, etc. If the ratings are determined based extensively on the statistics they would be more accurate.
- Every market participant could produce its own rating scale to classify the other participants. Thus, independent credit rating estimation could use its own data and the distributed overall assessment could increase the quality of the taken decisions.

2 SOLUTION

The stages of the proposed system are shown in fig. 1. The first stage is choosing the grades of the rating scale. Different agencies work with different scales which contain upper and lower alphabetical letters combined with positive or negative signs and sometimes digits. Some rating agencies use different scales for the long and short term rating. In our system only one scale is used. The information for the scale is the first needed information because it determines the number and order of the rating grades which define the model building step.

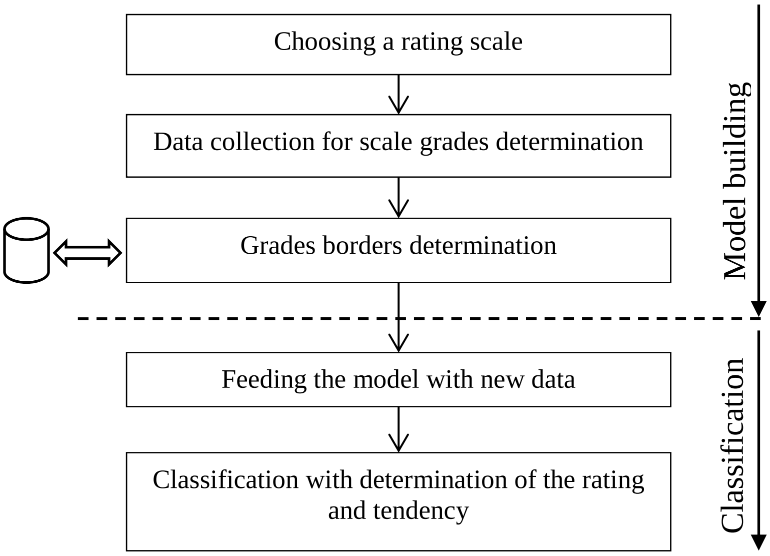

Fig. 1. The stages of implied rating system working

In the next step the available data should be collected in order to determine the centers and boundaries of the grades. In our approach this data comprises of the levels of Credit Default Swap (CDS) spread curves, Credit Default Swap Index (CDX) spread indices [1], share prices and bond prices which theirself contain information reflecting the current worthiness of the participants represented by such data. The working of the automated rating determination system relies on the reverse task of the upper data determination.

Thus by given data the credit rating of a participant should be calculated by mathematical approaches. The spread curves and indices could be considered similar to assurance which means that the higher their values are the higher the risk of the default is for the participant they represent. The data used for the scale building are in fact a set of time series for a given historical time period. Every time series represent a market participant. The time horizon should be the same for all series. If the values do not coincide by dates then interpolation is performed to make them for equal dates. The more data collected the more accurate rating scale will be build.

The grades centers and bounds determination is performed by a mathematical model composed by a self-organizing features map [7] with one-dimensional output lattice which is described in more details below. Once the scale is built it could be saved in a database for use after that. Periodically the scale must be built and saved again. Thus the whole scale with all grades fluctuates over time according to the state of all participants which series are used in the scale building. This means that if for example a crisis happens then the series of the most participants would move upward because the interest rates increase but this occurs correlated for them and thus most participants could also preserve their rating.

The scale building is based on ranges determination from the given data. In fact the series may overlap one another in some time periods but the grades boundaries are finally precisely specified. The degrees are determined by their centers. Having these centers the boundaries between them are calculated as their averages. When a new series is classified it could crosses some of the boundaries but this should not hinder from finding the nearest degree center to the series. The simple or decayed Euclidean distance is used in this case as it is shown to be d1 and d2 in fig. 2 where the new series center is shown in yellow and it is between the grades AA and A.

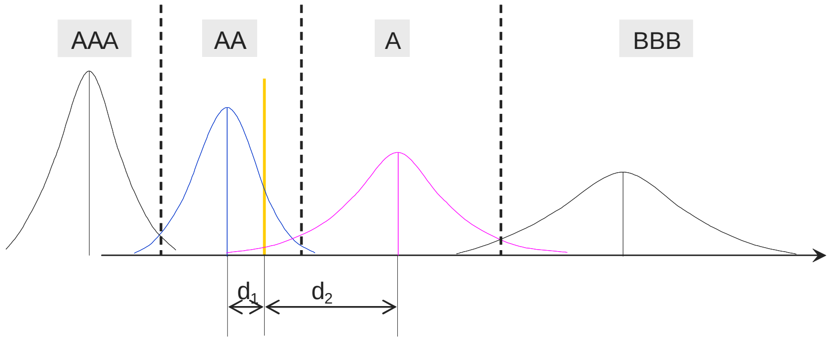

Fig. 2. Classification of a new obligator series by measuring the distance

All data values in every degree are considered as a set and their probability distribution are built supposing normal distribution. To do this first the normal distribution parameters μ (mean) and σ (standard deviation) are estimated for every degree and then the theoretical histogram is graphically shown [3] [4]. The new series together with its degree center and boundaries are shown in fig. 3.

3 THE MATHEMATICAL MODEL

3.1 Model building

The mathematical model used in our approach is based on self-organizing map used for clustering and classification of time series. Its input and output layers are shown in fig. 4. This self-learning mathematical model determines the parameters of its internal structure based on the input data. The topological ordering of the output nodes is one-dimensional in our case because the grades are ordered one-dimensionally. The spatial shape of an output unit is chosen to be a square. Every output node corresponds to a group which represents a rating scale degree. Each input node corresponds to a single historical time series value. Thus in the learning stage of the self-organizing map the time series are classified into groups based on both their magnitude and historical behaviour. The learning is an iterative process in which all time series are subsequently used as input and the output nodes prototypes change their positions. In the beginning of the learning the changes are greater and with the time it decays non-linearly doing fine correction until the end.

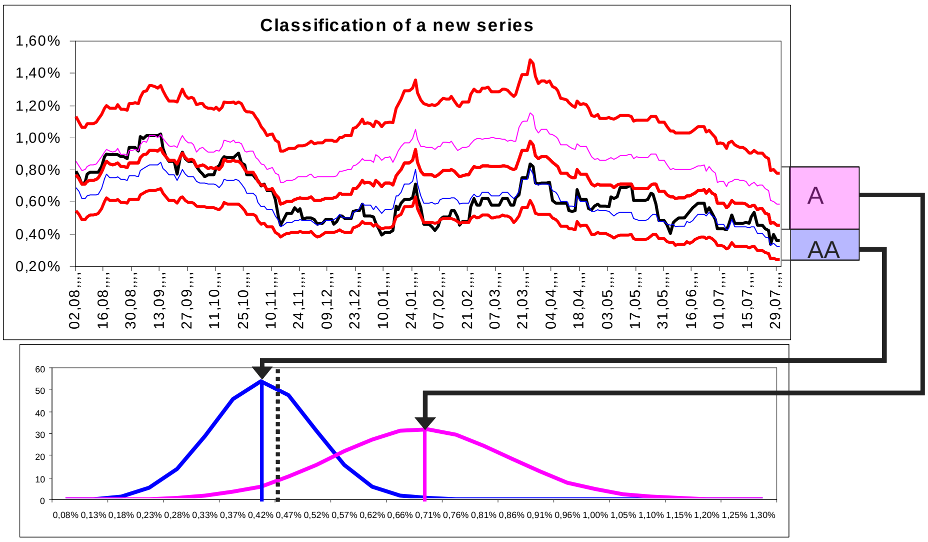

Fig. 3. A new series (in black color) shown together with its degree and the second nearest degree

When the groups and their centers are determined the grades bounds are calculated based on the groups' centers which correspond to the output nodes prototypes. The greater the rating is the smaller the prototype magnitude is. The prototypes are shown in figure 4 in the output layer nodes as small curves in the squares. The prototypes are considered as grades centers so groups' bounds are calculated to be the average of the neighbors' prototypes which can be seen in figure 3 where the center of the group representing the grade A is shown in pink and the center of AA in blue. Thus, if the width of two neighbor grades is not the same then the prototype will not be in the center of the grade after its bound determination.

The grades determination is followed by their reordering according to the mean of their prototypes. This is needed because the self-organizing map weights are initialized with random values and the grades must always be sorted in the same way. When the group determination should be repeatable the random generation is set appropriately. Thus the groups are sorted descending after the learning stage.

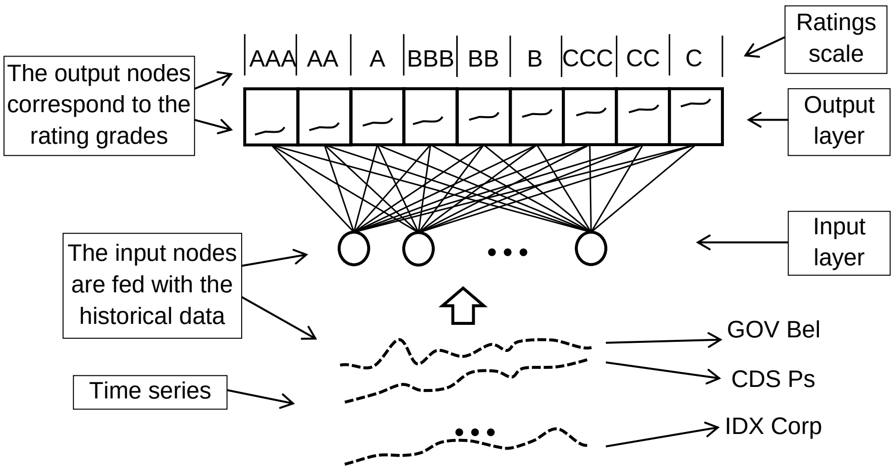

Fig. 4. The self-organizing map based model

3.2 Using the model for classification of a new series

The prototypes of the output nodes are used not only for the determination of the grades but also for classification in the next stage. When the rating scale is determined the next stage is to use it in order to classify a new series. In this way the rating category of a participant is determined. This is performed by comparing the new series with each prototype of output node and finding the nearest one which is chosen to be the rating grade. In the classification stage not only the best matching grade is found but also the second nearest one which is considered as the tendency grade with a given confidence. The confidence is calculated as closeness between the new input series and the prototypes of the rating (nearest) and tendency (second nearest) grades using (1) – (3).

$$s = r + t \tag{1}$$

$$v_r = 100 * \left(1 - \frac{r}{s}\right) \tag{2}$$

$$v_t = 100 * \left(1 - \frac{t}{s}\right) \tag{3}$$

where
r – distance between the time series and the nearest (rating) prototype;
t – distance between the time series and the second nearest (tendency) prototype;
v_r – confidence of the rating;
v_t – confidence of the tendency;

3.3 Using of decay factor

An important fact is that the model is built not only giving an account of the data magnitude but also of their historical behavior. The more recent data however should be considered as more important in both the grades determination and for the classification stage. That is why a decay factor is used when the distance is calculated between a data time series and a grade prototype using (4).

$$d = \sqrt{\frac{1}{\sum_{i=1}^{N} \lambda^{i-1}} \sum_{i=1}^{N} \lambda^{N-i} (x_i - y_i)^2} \tag{4}$$

where
λ – a decay factor;
N – time series size;
x – time series;
y – grade prototype;

The decay factor values vary from 0 to 1. When a decay factor is used in the grades determination stage the entire set of time series in the grade tend to be within the grades bounds at their end. In fig. 5 in the left a grade is shown with time series without using a decay factor and in the right the grade is shown using it. It can be seen that at the end of the X axis the series tend to be between the grade bounds only if the decay is used. Otherwise the series scatter also outside of the grade bounds through all their length.

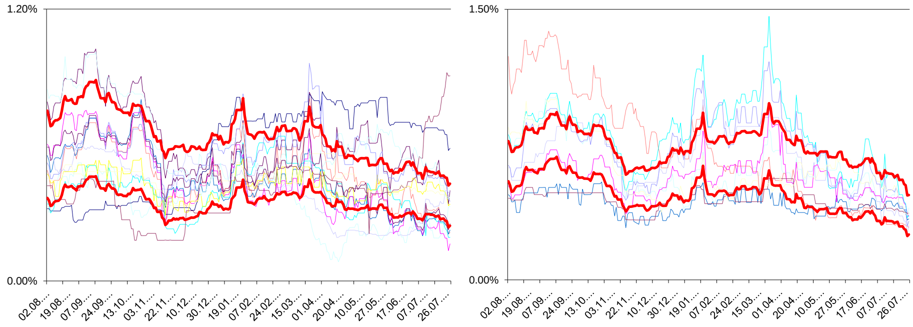

Fig. 5. The effect of the using of the decay factor during the grades determination stage

The decay factor could also be used for the mean calculation of a series. The effect can be seen in fig. 5 where the red dashed line is the mean with decay giving more weight to the last values in the series and with blue dashed line the mean is shown calculated without a decay factor.

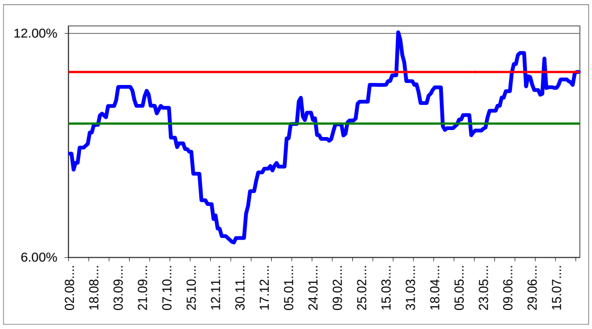

Fig. 6. Calculation of mean of a series with decay factor (dashed red line) and without it (dashed green line)

4 CONCLUSIONS AND FUTURE WORK

The prototype of the software system realizing the described approach is developed in Java. Its computational part is implemented as a JAR library which can be used either as a directly used module in business layer logic of other software systems or as service operations.

Below some examples of classification of new series are shown. The grades are ordered from AAA that is the best credit rating to D that is default. In fig. 7 the new series is classified as CC and despite of the fact that its trend is increasing its tendency is shown to be CCC because some part of the history has been in that grade.

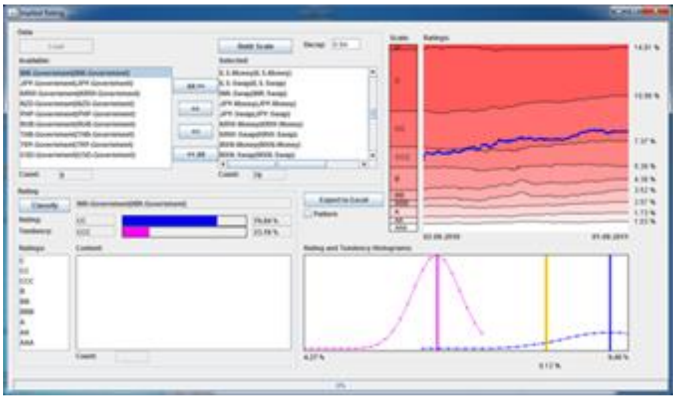

Fig. 7. The implemented prototype

In fig. 8 the series is classified as the best rating AAA with tendency to be AA. There could not be other tendency grade because there is no better credit rating than AAA. That is why in such cases the most important information is to what extend the tendency is to be AA.

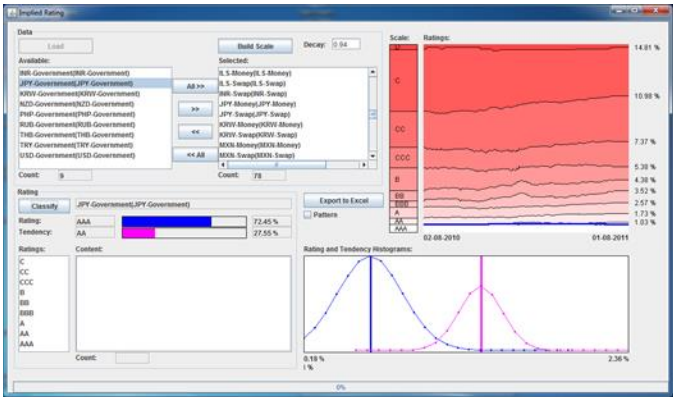

Fig. 8. The implemented prototype

In fig. 9 the series is classified to be BB but is moving too near to the board between BB and BBB. In the history it has been in the neighbour BBB grade.

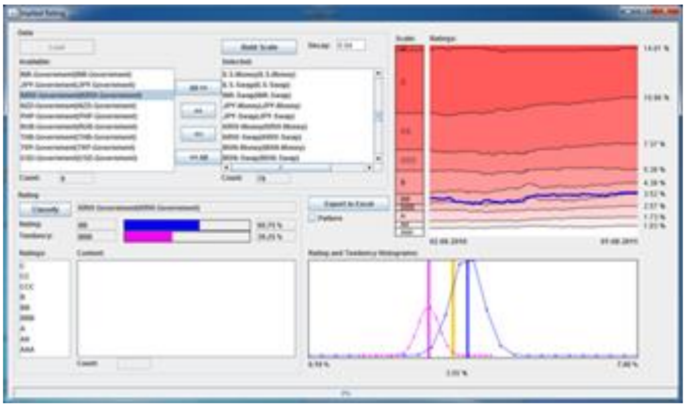

Fig. 9. The implemented prototype

An interesting series is shown in fig. 10 where the series starts from B moves in BB and finishes in BBB. The decay factor here shows the importance of the last values in such cases. Moreover, the tendency is shown to be A even though the series has been from the other side of the scale and never in A.

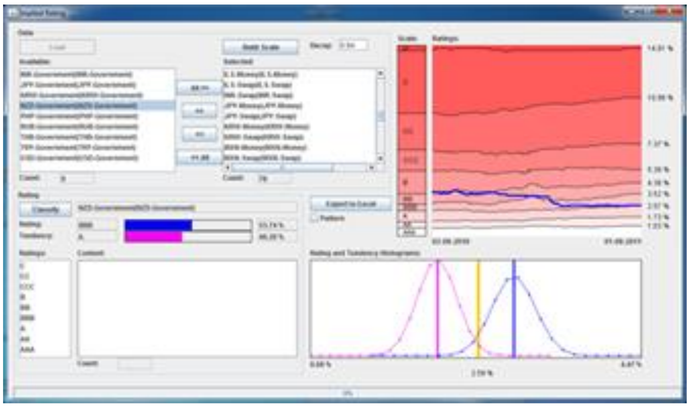

Fig. 10. The implemented prototype

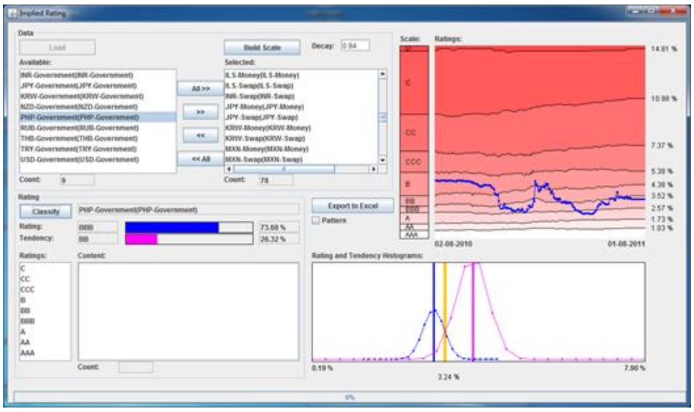

Fig. 11. The implemented prototype

Fig 11 also shows a case when the decay is important for the right classification. The series is determined to be in BBB thought in his historical movement it changes from B to A.

The experiments show that the system is robust especially regarding the ability to classify according to the more actual data. Thus the historical values are taken into account but not so important than the last ones. And such a system could be used on daily basis and with individual settings that are good advantages not only for experimental but also for practical uses.

5 REFERENCES

1. Beinstein, E. et al. Credit Derivatives Handbook Detailing credit default swap products, markets and trading strategies. Corporate Quantitative Research New York, London December, JPMorgan. 2006.
2. Standard& Poor's. Global Credit Portal – Standard and Poor's Rating Definitions. June 22, 2012.
3. Bury, K. Statistical Distributions in engineering. Cambridge University Press, 1999.
4. Krishnamoorthy, K. Handbook of statistical distributions with applications. Chapman & Hall, 2006.
5. Mun, J. Modeling Risk. Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization Techniques. Wiley, 2006.
6. Vallentin, M. Probability and Statistics Cookbook, 2011 (http://matthias.
7. vallentin.net/probability-and-statistics-cookbook/).
8. Kohonen, T. Self-organizing maps. Springer, 2001.

6 ABOUT THE AUTHOR

Dr. Ventsislav Nikolov
Senior Software Developer
Eurorisk Systems Ltd.
31, General Kiselov Str., 9002 Varna, Bulgaria
E-mail: vnikolov at euroriskssystems dot com