Department of Economics and Management

DEM Working Paper Series

Deep Learning Bank Distress from News and Numerical Financial Data

Paola Cerchiello
(Università di Pavia)

Giancarlo Nicola
(Università di Pavia)

Samuel Rönnqvist
(Åbo Akademi University)

Peter Sarlin
(Hanken School of Economics, RiskLab Finland)

 # 140 (05-17)

May 2017

Deep learning bank distress from news and numerical financial data

P. Cerchiello[†], G. Nicola[†1] S. Rönnqvist[*] and P. Sarlin[+‡]

[†] Dep. of Economics and Management Science, University of Pavia.
[*] Turku Centre for Computer Science – TUCS, Åbo Akademi University, Turku, Finland.
[+] Department of Economics, Hanken School of Economics, Helsinki, Finland.
[‡] RiskLab Finland, Arcada, Helsinki, Finland.

Abstract

In this paper we focus our attention on the exploitation of the information contained in financial news to enhance the performance of a classifier of bank distress. Such information should be analyzed and inserted into the predictive model in the most efficient way and this task deals with all the issues related to text analysis and specifically analysis of news media.

Among the different models proposed for such purpose, we investigate one of the possible deep learning approaches, based on a doc2vec representation of the textual data, a kind of neural network able to map the sequential and symbolic text input onto a reduced latent semantic space. Afterwards, a second supervised neural network is trained combining news data with standard financial figures to classify banks whether in distressed or tranquil states, based on a small set of known distress events. Then the final aim is not only the improvement of the predictive performance of the classifier but also to assess the importance of news data in the classification process. Does news data really bring more useful information not contained in standard financial variables? Our results seem to confirm such hypothesis.

JEL classification: C83, C12, E58, E61,G02, G14.
Keywords: Behavioural Finance, Financial News, Deep Learning, Bank Distress, Word2vec.

[1] Corresponding Author
giancarlo.nicola@universitadipavia.it
[-2pt] This version April 2017

1 Introduction

The interpretation of text by machines, the task of natural language processing, is complex due to the richness of human language, as well as the ambiguity present at many levels, including the syntactic and semantic one. From a computational point of view, processing language means dealing with sequential, highly variable and sparse symbolic data, with surface forms that cover the deeper structures of meaning. Despite these difficulties, there are several methods available today that allow for the extraction of part of the information content present in texts. Some of these rely on hand-crafted features, while others are highly data-driven and exploit statistical regularities in language. Among the statistical methods, many rely on word representations. Class based models, for example, learn classes of similar words based on distributional information, like Brown clustering [1] and Exchange clustering [2],[3]. Soft clustering methods, like Latent Semantic Analysis (LSA) [4] and Latent Dirichlet Allocation [5], associate words to topics through a distribution over words of how likely each word is in each cluster. In the last years many contributions employ machine learning and semantic vector representations [6],[7], lately using Long Short-Term Memory (LSTM) networks [8],[9],[10] to model complex and non-local relationships in the sequential symbolic input, Recursive Neural Tensor Networks (RNTN) for semantic compositionality[11][12] and also convolutional networks (CNN) for both sentiment analysis [13] and sentence modelling [14]. In this vein, Mikolov et al. [6][15][27] have introduced an unsupervised learning method to create a dense multidimensional space where words are represented by vectors. The position of such vectors is related to their semantic meaning, further developing the work on word embeddings [16] which grounds on the idea of distributed representations for symbols [17]. The word embeddings are widely used in modern NLP since they allow for a dimensionality reduction compared to a traditional sparse vector space model.

Another approach to improve information extraction from text is to enhance the algorithm with contextual information related to the current environment in which the system is operating. Yet, the introduction of contextual information in the models is not a straightforward process but requires a careful choice of the additional information provided in order to not increase noise. These advancements in text analytics aim at increasing the potential value of text as a source in data analysis [31]. In our case, the application we focus on consists in understanding bank distress, an area where text data hold promising potential due to the frequency and information richness of financial news. In fact, recently, central banks are starting to recognize the utility of text data in financial risk analytics [18][19].

This recent rise of interest around text-based computational methods for measuring financial risk and distress is fuelling a rapidly growing literature [30]. The most covered area is sentiment analysis to correlate it with events of interest. Many of the previous approaches have been based on hand-crafted dictionaries that despite requiring work to be adapted to single tasks can guarantee good results due to the direct link to human emotions and the capability of generalizing well trough different datasets. Examples of this kind are the works of Nyman et al. [20] and Soo [21]. The first analyzes sentiment trends in news narratives in terms of excitement/anxiety and find increased consensus to reflect pre-crisis market exuberance, while the second correlates the sentiment in news with the housing market. Despite the good results, there are applications where it could be preferable to avoid dictionaries in favour of more data driven methods, which have the advantage of higher data coverage and capability of going beyond single word sentiment expression. Malo et al. [22] provide an example of a more sophisticated supervised corpus-based approach, in which they apply a framework modelling financial sentiment expressions by a custom data set of annotated phrases. As Rönnqvist & Sarlin [23] applies a fully data driven model with unsupervised semantic generalization, supervised only by a small set

of events, this paper extends upon that work by incorporating numerical financial data as well.

This study aims at demonstrating the feasibility and usefulness of the integration of textual and numerical data in a machine learning framework for financial predictions. Thus, the goal of the predictive model is to correctly classify stressed banks from both financial news and financial numerical data. In Section 2 we illustrate the data and the predictive task, in Section 3 we describe the machine learning framework, in Section 4 we present the experimental results with a sensitivity analysis on the network parameters and in Section 5 we show the conclusions of the work with hints on future developments.

2 The Data

In this paper we leverage two types of data, textual and numerical, matched by chronology and entities, to predict event data. The event dataset contains information on dates and names of involved entities, relating to the specific type of event to be modelled, while the textual and financial numerical datasets contain respectively bank related articles and financial figures. The textual and numerical data are connected to a particular event using date and occurrences of the entity name for the textual data and matching the date with the corresponding quarter for the financial numerical data. The model is then trained in a supervised framework to associate specific language and financial figures with the target event type.

2.1 Textual and numerical data

The textual data are taken from a database of news articles from Reuters online archive spanning the years from 2007-Q1 to 2014-Q3. The original data set includes 6.6M articles, for a total of ca. 3.4B words. In order to select only articles related to the banks object of study, as in [23], we have looked at bank name occurrences and selected only the articles with at least one occurrence. Bank name occurrences are located using a set of patterns defined as regular expressions that cover common spelling variations and abbreviations of the bank names. The regular expressions have been iteratively developed against the data to increase accuracy, with a strong attention on avoiding false positives. As a result from the entire corpus we retrieve 262k articles mentioning any of the 101 target banks. Successively the articles are split in sentences and only the sentences with bank name occurrences are kept. We integrate contextual information through a database of distress related indicators for banks. The numerical data is composed of 12 variables for 101 banks over the period 2007-2014 with quarterly frequency. In table 1 we report the list of numerical variables, among them there are information on bank-level balance sheet and income statement data, as well as country-level banking sector and macro-financial data (table 1).

Bank Level	Bank Sector Level	Macro Level
Capital to asset	Mortgages to loans d4	House price gap (Deviation from trend of the real residential property price index)
Interest to liabilities	Securities to liabilities d4	Macroeconomic Imbalance Procedure (MIP), international investment position
Reserves to asset	Financial assets to gdp	Private debt
-	-	Government bond yeld
-	-	Credit to gdp
-	-	Credit to gdp delayed 12 months

Table 1: List of available numerical variables.

In Table 2 we report summary statistics of the analyzed numerical variables.

Then, we match the distress events with the available textual news data. The events are based upon bankruptcies and direct defaults, government aid and distressed mergers as presented in [24]. The

2

Variable	Mean	Variance	Standard Deviation	Kurtosis
Capital to asset	2.5	10.2	3.2	21.5
Reserves to asset	4.2	8.5	2.9	4.3
Interest to liab	3.4	8.5	2.9	104.6
Financial assets to gdp	385.0	134,365.2	366.6	33.2
Mortgages to loans d4	0.2	1.7	1.3	0.1
Securities to liab d4	-12.0	1,342,234.9	1,158.5	105.7
Credit to gdp	140.2	2,623.4	51.2	0.0
Credit to gdp d12	13.7	479.1	21.9	0.5
House Price Index rt16 gap	-2.5	33.7	5.8	6.8
International Investment Position	-21.0	2,967.7	54.5	0.0
Private Debt	188.2	4938.7	70.3	0.1
Gov Bold Yield d4	0.0	11.4	3.4	23.5

Table 2: Summary statistics of available numerical variables.

distress events in this dataset are of three types. The first type of events include bankruptcies, liquidations and defaults, with the aim of capturing direct bank failures. The second type of events comprise the use of state support to identify banks in distress. The third type of events are forced mergers, which capture private sector solutions to bank distress. The inclusion of state interventions and forced mergers is important to better represent bank distress since there have been few European direct bank failures in the considered period. Bankruptcies occur if a bank net worth falls below the country-specific guidelines, whereas liquidations occur if a bank is sold and the shareholders do not receive full payment for their ownership. Defaults occur if a bank has failed to pay interest or principal on at least one financial obligation beyond any grace period specified by the terms or if a bank completes a distressed exchange. The distress events are defined to start when a failure is announced and end at the time of the de facto failure.

A capital injection by the state or participation in asset relief programs (i.e., asset protection or asset guarantees) is an indication bank distress. From this 'indicator' are excluded liquidity support and guarantees on banks liabilities since they are not used for defining distressed banks. The starting dates of the events refer to the announcement of the state aid and the end date to the execution of the state support program. Distressed mergers are defined to occur if (i) a parent receives state aid within 12 months after a merger or (ii) if a merged entity exhibits a negative coverage ratio within 12 months before the merger. The dates for these two types of distress events are defined as follows, respectively: (i) the starting date is when the merger occurs and the end date when the parent bank receives state aid, and (ii) the start date is when the coverage ratio falls below 0 (within 12 months before the merger) and the end date when the merger occurs.

2.2 Data Integration

The following step in the data preparation has been the addition of numerical financial data to the text news database. The numerical data was aligned with the set of sentences in which bank names occurred. The purpose was to match each and every mention of a bank with the corresponding numerical financial data aligned according to the same time horizon. Since the news and the financial data have different frequency, in particular, news have higher frequency while financial data are reported quarterly, the latter are replicated several times to perfectly match with the former. For each news regarding a bank within a given quarter, financial data are replicated and appended to the semantic vector of the news. Such matching activity between numerical and textual data resulted in the removal of a few banks from the dataset due to missing data, causing a reduction from 101 to 62 target institutes (Table 3) and from about 601k to 380k news sentences. After cleaning the dataset, numerical data have been normalized to improve the classifier training time, due to easier convergence of the model. We have normalized the data with a standard approach, that is, by subtracting the mean value from each numerical variable of the dataset and dividing it by the standard deviation. The

resulting input vector for the 612 dimensional input layer of the neural classifier is thus composed of a numerical vector obtained joining together the 600-dimensional semantic vector coming from the unsupervised modelling, described in section 3, with the 12-dimensional numerical financial data vector. The dataset is then split into five folds, three for training, one for validation and one for testing according to a cross-validation scheme. The folds are created so that all the data regarding a given bank are in the same fold. The framework we apply is composed of an unsupervised algorithm and a supervised neural network classifier. To train the classifier we need to provide a label indicating the distressed or tranquil status of the bank. The dataset has been labelled according to the bank status with 0 for tranquil and 1 for distress. The proportions of the two classes fed to the network are highly unbalanced: 93% of the data-points represent tranquil status and only the remaining 7% associated to distress events. Such imbalance of the classes has a significant impact both on the training and on the evaluation of the model. Regarding the training, it is important that the model is able to generalize due to the few distress examples, while for the evaluation it is important to provide an alternative measure to the accuracy. Using this method, in fact, a trivial model that always predicts the tranquil status would achieve a 93% accuracy. Thus, it is necessary to measure an improvement against this baseline. Moreover the user is likely interested in weighting differently first error and second error types, especially in early warning applications. The usefulness measure, introduced in [25], is able to handle these requirements.

Financial Institution	Country	Sector	Financial Institution	Country	Sector
Aareal Bank	DE	Carnegie Investment Bank	SE	Kommunalkredit	AT
ABN Amro	NL	Commerzbank	DE	LBBW	DE
Agricultural Bank of Greece	GR	Credit Mutuel	FR	Lloyds TSB	UK
Allied Irish Banks	IE	Credito Valtellinese	IT	Max Bank	DK
Alpha Bank	GR	Cyprus Popular	CY	Monte dei Paschi di Siena	IT
Amagerbanken	DK	Danske Bank	DK	National Bank of Greece	GR
ATE Bank	GR	Dexia	FR	Nordea	SE
Attica Bank	GR	EBH	DK	NordLB	DE
Banca Popolare di Milano	IT	EFG Eurobank	GR	Nova ljubljanska banka Group (NLB)	SI
Banco Popolare	IT	Erste Bank	HU	OTP Bank Nyrt	HU
Bank of Cyprus Public Co Ltd	CY	Fionia (Nova Bank)	DK	Piraeus Bank	GR
Bank of Ireland	IE	Fortis Bank	B	Pronton Bank	GR
Banque Populaire	FR	HBOS	UK	RBS	UK
Bawag	AT	Hellenic	GR	Roskilde Bank	DK
BayernLB	DE	HSH Nordbank	DE	Societe Generale	FR
BBK	BRN KWT	Hypo Real Estate	DE	Swedbank	SE
BNP Paribas	FR	Hypo Tirol Bank	AT	T-Bank	GR
BPCE	FR	IKB	DE	UNNIM	ES
Caixa General de Depositos	P	ING	NL	Vestjysk	DK
Caja Castilla-La Mancha	ES	Irish Nationwide Building Society	IE		
CAM	ES	KBC	BE		

Table 3: List of considered financial institutions

3 Deep learning framework

Machine learning systems benefit from their ability to learn abstract representations of data, inferring feature representations directly from data instead of relying on manual feature engineering. This capability is particularly exploited in deep learning models, which provides flexibility and potentially better accuracy [26]. In particular the flexibility is very suitable for natural language processing tasks where the ability to generalize across languages, domains and tasks enhances the applicability and robustness of the text analysis. The framework applied in this paper is an extension of the one developed in [23] with the aim of predicting bank distress from textual data. Their scheme allows to infer banks distress conditions from textual news with a machine learning approach based on two steps:

- The first step applies an unsupervised algorithm to retrieve the semantic vectors associated to a text. Using the Distributed Memory Model of Paragraph Vectors (PV-DM) by Le & Mikolov [27] (here referred to as doc2vec) for learning vector representations of sequences of words, dense vector representations of sentences mentioning target banks are learned. This method creates a multidimensional space (600 dimensions), where words are positioned according to

their semantic meaning (the closer the meaning of two words, the closer are their positions). From this new space is easier to perform the classification task due to the reduced dimensionality and the wise positioning of the vectors that takes into account their semantic meaning.

- The second step performs the classification task through a supervised algorithm. The sentence vectors are fed into a neural network classifier. The neural network is composed of one input layer (600 nodes), one hidden layer (50 nodes) and one output layer (2 nodes with stress prediction $e \in \{0, 1\}$).

In this paper we modify the previous machine learning setup to integrate the support of financial numerical data. The financial data that we integrate hold information about bank accounting data, banking sector data and country macroeconomic data. The two-step structure of the data processing flow has been preserved.

The approach here used to learn the semantic vectors is a Distributed Memory Model Paragraph Vector [27]. In this model the semantic vector representation is learned by training a feed forward neural network to predict the successive word based on a word context (previous n words) and the randomly initialized semantic vector (sentence ID). The contexts to predict the next words are fixed-length and sampled from a sliding window over the sentence. While training the network the semantic vector gets updated by the training algorithm so that its representation positively contributes in predicting the next word and thus works as a semantic representation of the entire sentence. In this way the sentence ID works as a memory for the model that allows the vector to capture the semantics of continuous sequences; the sentence ID in fact can be thought of as an extra word representing the sentence as global context and conditioning the prediction of the next word. Despite the fact that semantic vectors are initialized randomly they gradually improve how they capture the semantic of the sentence during the training, performed by stochastic gradient descent and with the gradient computed via backpropagation algorithm. Formally, the training procedure seeks to maximize the average log probability:

$$\frac{1}{t+n} \sum_{i=1}^{t-n} \log p(w_{i+n+1} | s, w_i, ..., w_{i+n}) \tag{1}$$

over the sequence of training words $w_1, w_2, ..., w_t$ in sentence s with word context size n. The dimensionality of the semantic vectors and the context size of the algorithm have been optimized by cross-validation.

The second step, performing the classification task, receives in input two different sources: news textual data on the banks, in form of multidimensional semantic vectors V_s, and numerical financial data F_s loaded from a database matched with news data. For the classification task we employ a three layers fully connected feed forward neural network. The neural network has a 612 dimensional input layer (600 input nodes for the semantic vector V_s dimensionality and 12 input nodes for the numerical data F_s dimensionality), 50 hidden nodes and 2 output nodes for $e \in \{0, 1\}$ encoding the distress and tranquil status in a softmax layer that applies a cross-entropy loss function (see fig. 3.1). Finally, an additional phase of cleaning and merging of textual and numerical data has been added. The network is trained by Nesterovs Accelerated Gradient [29] to predict distress events $e \in \{0, 1\}$. Hence, the objective is to maximize the average log probability:

$$\frac{1}{|S|} \sum_{s \in S} \log p(e_s | V_s, F_s) \tag{2}$$

In the trained network, the posterior probability $p(e_s = 1 | V_s, F_s)$ reflects the relevance of sentence s to the modeled event type.
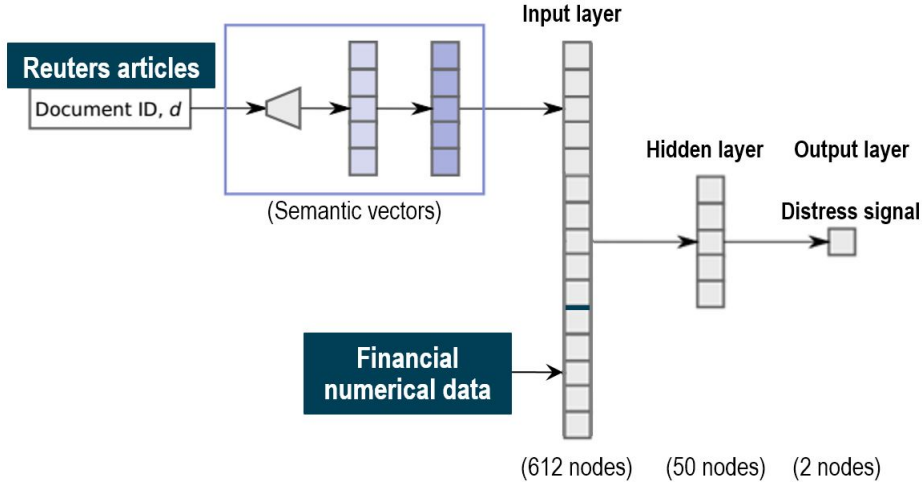
Figure 3.1: Structure of the model

4 Results

The experimental results confirm that the integration of numerical and textual data amplifies the prediction capability of the model compared with textual data alone. The distress events in the database represent only 7% of the cases, resulting in very skewed training classes as explained earlier. Moreover, given the nature of the problem, the identification of distress situations, it could be useful to weight differently false positives and false negatives. In an early warning application, a sensitive system is often preferable since a further investigation phase follows the detection of possible events. These peculiarities have to be taken into account during the evaluation of the model.

4.1 Evaluation and experimental results

To manage the peculiarities of the evaluation of our particular problem we resort to the relative usefulness as measure of performance. The relative usefulness (U_r), introduced in [25] is a measure that allows to set the error type preference (μ) and to measure the relative performance gain of the model over the baseline compared to a perfect model. Such index is based on the appropriate combination of the probabilities of the true positive (TP), false positive (FP), true negative (TN) and false negative (FN) that generate the model loss L_m (eq. 4) and on a baseline loss L_b set to be the best guess according to prior probabilities $p(obs)$ and error preferences μ (eq. 3).

$$L_b = min \begin{cases} \mu * p(obs = 1) \\ (1 - \mu) * p(obs = 0) \end{cases} \qquad (3)$$

$$L_m = \mu * p(FN) + (1 - \mu) * p(FP) \qquad (4)$$

The absolute Usefulness (U_a) and the relative Usefulness (U_r) are directly derived from the loss functions:

$$U_r = \frac{U_a}{L_b} = \frac{L_b - L_m}{L_b} \qquad (5)$$

As we can see from eq. 5 the relative usefulness is equal to 1 when the model loss (L_m) is equal to 0, thus the model is perfect. As a consequence, the relative usefulness measures the gain over the

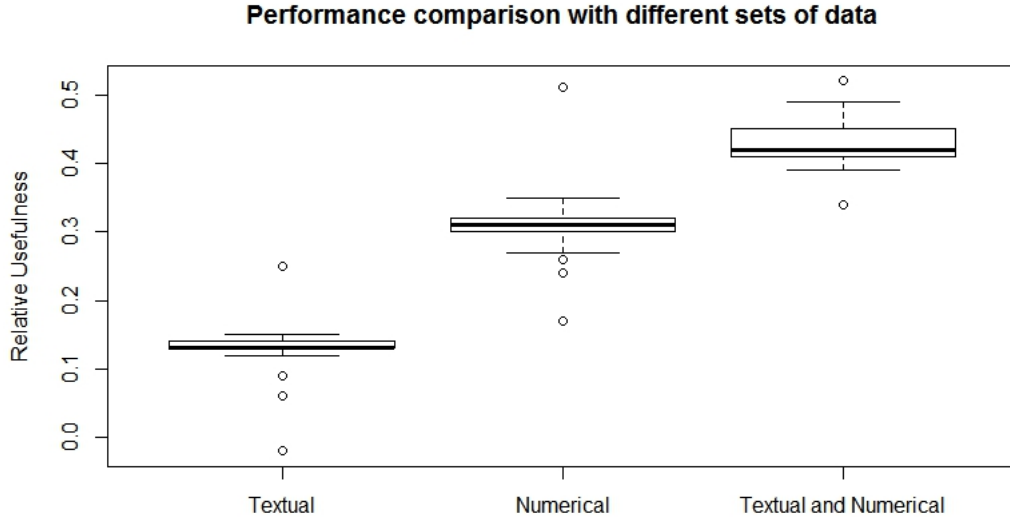**Performance comparison with different sets of data**

Figure 4.1: Comparison of the relative usefulness obtained with the textual financial data (left), numerical financial dataset (center) and with their combination (right)

baseline compared to the gain that an ideal model would achieve. To compute the relative usefulness (U_r) we have set the error type preference (μ) equal to 0.9 in accordance with the indications of previous studies [24][28] on the importance of signalling every possible crisis at cost of some false positive (FP) (with $\mu = 0.9$ we are saying that missing a crisis is about 9 times worse than falsely signalling one). This is especially true if following the signalling of an event, a further investigation action is triggered. In order to evaluate distress condition of a bank over a period, the predictions are aggregated on a monthly basis by bank entity. This is done averaging the predictions at the single sentence level by month for each different bank. This has been done to take into account the information available over one month period reducing the predictions variability. As a result of this procedure, the classification task can be summarized as understanding which banks are in distress conditions month by month based on the news and numerical data available over the previous month.

To evaluate the model on this classification task, we have trained it fifty times on the same dataset, recording the relative usefulness (U_r) result after each run and then averaging on them. For each of the fifty trainings, the folds are resampled and the neural net is randomly initialized. To quantify the gain obtained from merging numerical and textual data we have done three different experiments, running the model respectively with textual data only (fig. 4.1, left), numerical data only (fig. 4.1, center) and numerical and textual data together (fig. 4.1, right). As it is possible to see in fig. 4.1 the case with textual data alone achieves an average relative usefulness of 13.0%, while the case with numerical data alone shows an average relative usefulness of 31.1%. The combination of these two dataset and their exploitation in the model grants an average relative usefulness of 43.2%, thus it positively enhances the prediction capability of the model. From these results we can also understand that, as expected, the financial numerical data hold the majority of the informative potential necessary for the labelling task but that the addition of textual information provides a non-negligible 12.1% improvement to the relative usefulness of the model.

4.2 Classifier tuning

In order to improve the classifier performance we have run a sensitivity analysis exploring different neural network configurations while training it with the Nesterov Accelerated Gradient Descent al-

gorithm [29]. We have tested different hidden layers dimensionalities, numbers of layers, learning rates, regularization parameters and dropout fractions [31]. For choosing the network configuration, we apply the Occams razor principle always preferring the simpler structure able to achieve a given performance. Thus, where performance is not reduced excessively, we try to select the network structure with fewer layers and fewer hidden nodes; this also helps to have better generalizing model and avoid overfitting when applying it to other datasets. In terms of hidden layers number, the network with one hidden layer (three layers in total including input and output) performs slightly better than those with more layers. We tested up to three hidden layers (5 layers in total) and the performance was monotonically decreasing. Regarding the number of hidden nodes, the network that gave the best usefulness has 50 hidden nodes with a learning rate α of 0.0005 combined with an L_1 regularization parameter λ of 0.00001. The parameter that mostly affects the results is the number of nodes in the hidden layer. The results of the sensitivity analysis on the hidden nodes number (with regards to one hidden layer network configuration) are reported in fig. 4.2, 4.3 and 4.4 respectively for the case including textual data alone, financial numerical data alone and the combination of the two. The range of hidden nodes in the three sensitivities is different because the input vectors in the three cases have very different dimensionalities, 600 input nodes when considering only textual data, 12 input nodes when considering only numerical data and 612 input nodes when including both numerical and textual data. We do not investigate extensively the textual data case which has already been studied in [23]. Regarding the numerical based case, we can notice that we have a range of hidden layer size comprised between 10 and 20 nodes where performances are stable and Relative Usefulness is around 30%. For the combined dataset (Numerical and Textual) we can see that there is a range around 50-60 hidden nodes where performance is stable around a 40% Relative Usefulness. We expected the right number of hidden nodes to be similar to the Textual data case since the input dimensionality is similar (600 and 612).
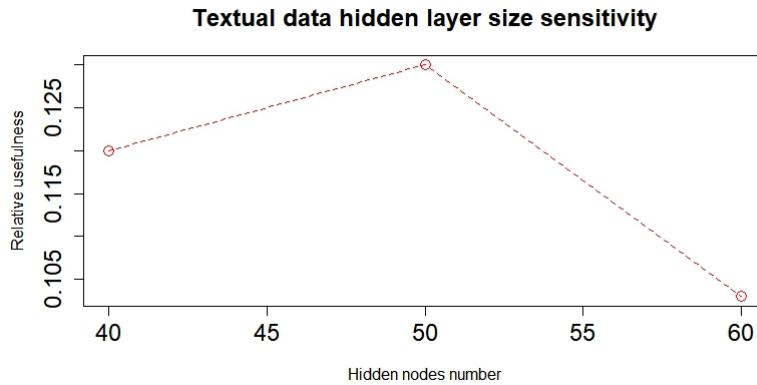

Figure 4.2: Textual data - sensitivity analysis on the number of nodes of the hidden layer
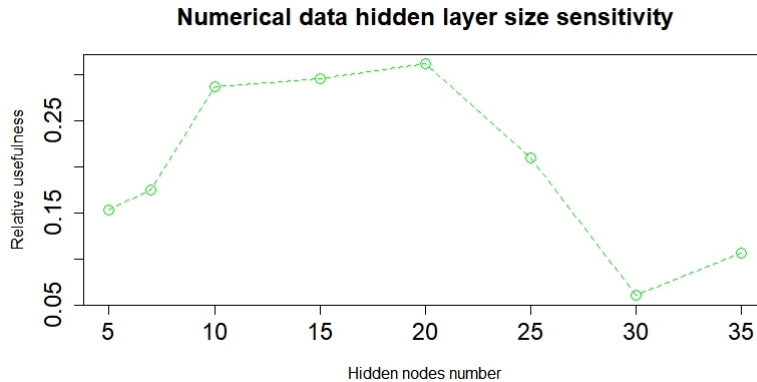

Figure 4.3: Numerical data - sensitivity analysis on the number of nodes of the hidden layer
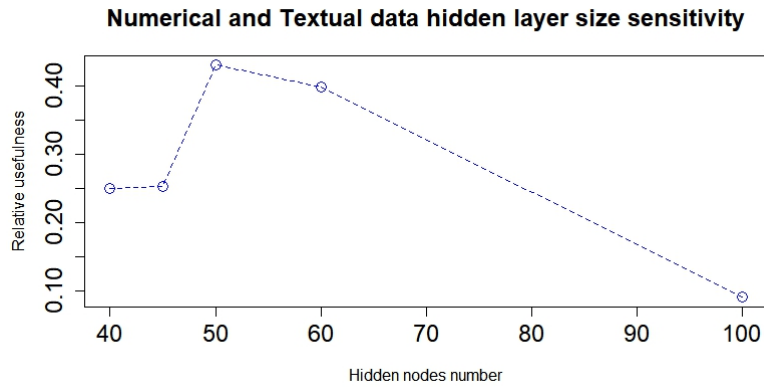
8

Numerical and Textual data hidden layer size sensitivity

Figure 4.4: Numerical and Textual data - sensitivity analysis on the number of nodes of the hidden layer

5 Concluding Remarks

In this work we have presented an approach for the integration of financial numerical data and financial news textual data into a single machine learning framework. The aim is to identify bank distress conditions with improved performance with respect to the model based on news data only. The implemented framework processes textual data through an unsupervised neural network model converting the documents sentences into sentence vectors. The retrieved sentence vectors can then be joined with the financial numerical data in a unique input vector and fed to a supervised classifier, in our case a three layers fully connected neural network.

The classification task was characterized by a high imbalance in the classes which poses concerns for both the training and the evaluation of the model. However, the implemented neural network is able to learn the combinations of the financial conditions of the banks and the semantic content of the news that are more associated with distress conditions. This is reflected in the improved performance obtained when leveraging both textual news data and financial numerical data, average relative usefulness of 43.2%, compared to 31.1% when using only numerical data and 13.0% when utilizing only textual data. The methodology here applied is general and extensible to other problems were the integration of text and numerical covariates can improve classification and early warning performances. Other interesting results are to be expected in areas where textual data hold information with higher granularity and frequency, directly influencing the data to be predicted in the short run like in the case of financial markets.

References

[1] Brown, P. F., Pietra, V. J. D., deSouza, P. V., Lai, J. C., and Mercer, R. L.. Class-based n-gram models of natural language. Computational Linguistics, 18(4):467479, 1992.

[2] Martin, S., Liermann, J., and Ney, H.. Algorithms for bigram and trigram word clustering. Speech Communication, 24, 1937, 1998.

[3] A. Clark. Combining distributional and morphological information for part of speech induction. In Proc. of EACL, 2003.

[4] Landauer, T.; Foltz, P. W.; Laham, D.. Introduction to Latent Semantic Analysis. Discourse Processes. 25 (23): 259284, 1998. doi:10.1080/01638539809545028

[5] D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent Dirichlet allocation. JMLR, 3:9931022, 2003.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. In Proceedings of Workshop at International Conference on Learning Representations, 2013.

[7] J. Pennington, R. Socher, C. D. Manning. GloVe: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 15321543, October 25-29, 2014, Doha, Qatar.

[8] S. Hochreiter, J. Schmidhuber. Long Short-Term Memory. Neural Computation 9(8):1735-1780, 1997

[9] K. Cho, B. van Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In EMNLP, 2014.

[10] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In EMNLP 2013, pages 16311642.

[11] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and Christopher D. Manning. 2011. Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP).

[12] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and Christopher Potts. 2013b. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 16311642, Stroudsburg, PA, October. Association for Computational Linguistics.

[13] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu and P. Kuksa. Natural Language Processing (Almost) from Scratch. Journal of Machine Learning Research, 12:2493- 2537, 2011.

[14] N. Kalchbrenner, E. Grefenstette, P. Blunsom. 2014. A Convolutional Neural Network for Modelling Sentences. In Proceedings of ACL 2014.

[15] T. Mikolov. Statistical Language Models Based on Neural Networks. PhD thesis, PhD Thesis, Brno University of Technology, 2012.

[16] Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin. A neural probabilistic language model. The Journal of Machine Learning Research, 3:11371155, 2003.

[17] Hinton, G. E., McClelland, J. L., and Rumelhart, D. E. Distributed representations. In Rumelhart, D. E. and McClelland, J. L., editors, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations, MIT Press, Cambridge, MA. pp 77-109.

[18] D. Bholat, S. Hansen, P. Santos, and C. Schonhardt-Bailey. Text mining for central banks. In Centre for Central Banking Studies Handbook, volume 33. Bank of England, 2015.

[19] J. Hokkanen, T. Jacobson, C. Skingsley, and M. Tibblin. The Riksbanks future information supply in light of Big Data. In Economic Commentaries, volume 17. Sveriges Riksbank, 2015.

[20] R. Nyman, D. Gregory, K. Kapadia, P. Ormerod, D. Tuckett, and R. Smith. News and narratives in financial systems: exploiting big data for systemic risk assessment. BoE, mimeo, 2015.

[21] C. K. Soo. Quantifying animal spirits: news media and sentiment in the housing market. Ross School of Business Paper No. 1200, 2013.

[22] P. Malo, A. Sinha, P. Korhonen, J. Wallenius, and P. Takala. Good debt or bad debt: Detecting semantic orientations in economic texts. Journal of the Association for Information Science and Technology, 65(4):782796, 2014.

[23] S. Rönnqvist, P. Sarlin. Bank distress in the news: Describing events through deep learning. Forthcoming in Neurocomputing 2017.

[24] F. Betz, S. Opric, T. A. Peltonen, and P. Sarlin. Predicting distress in European banks. Journal of Banking & Finance, 45:225241, 2014.

[25] P. Sarlin. On policymakers loss functions and the evaluation of early warning systems. Economics Letters, 119(1):17, 2013.

[26] J. Schmidhuber. Deep learning in neural networks: An overview. Neural Networks, 61:85117, 2015.

[27] Q. Le, T. Mikolov. Distributed Representations of Sentences and Documents. Proceedings of the 31 st International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP, volume 32.

[28] A. Constantin, T. Peltonen, P. Sarlin. Network linkages to predict bank distress. Journal of Financial Stability forthcoming (2017).

[29] Y. Nesterov. A method of solving a convex programming problem with convergence rate o (1/k2). In Soviet Mathematics Doklady, volume 27, pages 372376, 1983.

[30] P. Cerchiello, P. Giudici and G. Nicola. 2017. Twitter data models for bank risk contagion. Forthcoming in Neurocomputing.

[31] G. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. Salakhutdinov. 2012. Improving neural networks by preventing co-adaptation of feature detectors. CoRR, abs/1207.0580.