

# Future Graduate Salaries Prediction Model Based On Recurrent Neural Network

Jakub Siłka, Michał Wieczorek and Marcin Woźniak

*Faculty of Applied Mathematics*

*Silesian University of Technology*

*Kaszubska 23, 44-100 Gliwice, POLAND*

kubasilka@gmail.com, michal\_wieczorek@hotmail.com, marcin.wozniak@polsl.pl

**Abstract**—Prediction models are widely applied in several fields. In this study we present a discussion on using Recurrent Neural Network as predictor for salaries of future graduates. The model is based on feature analysis which leads to input values of the predictor. We have analyzed several compositions and ideas. As a result we have selected Recurrent Neural Network to be the most accurate. Presented results confirm this selection and show high precision.

**Index Terms**—Recurrent Neural Networks, Prediction model.

## I. INTRODUCTION

PREDICTION models are used in various fields to help on future trends estimation. We can find many places where analysis of potential outcomes for the future is assumed as a key factor. One of such topics is salary prediction. We can find many discussions and proposed models which analyse financial factors influencing level of the financial outcome for various professions, countries, ages, etc. Our model is based on Recurrent Neural Network to precisely fit the prediction by the applied structure and proposed modified gradient descent training algorithm. Results, compared for various methods, show that our proposition has much advance.

In [1] was presented an economical approach with considering social aspects to analyze salary ranging from higher education level. The model considered many features which may result in future fluctuation of the salary. In [2] the aspects of salary level related to academic roots was considered for european universities. Authors compared several of them to select the best possible fields of studies in european society. Another aspects widely considered in research on salary fluctuation is race. As presented in [3] the origin of workers may have an active influence on the level of their salary. Similarly, differences in the salary may be visible if we compare woman and man, a study in this field for academic libraries workers was presented in [4]. There are also several aspects which should be analyzed for different countries. In [5] we can read about historical analysis over work positions in India. There are many methods which are used in such studies, ie. [6] proposed deep learning approach model with Convolutional Neural Networks. In all analysis we can find social aspects, economic features and mathematical models. The best option is to model such approach with the most flexible but at the same time precise approach.

In our study we have decided to use methods of machine learning. There are several models which found application to prediction systems. However recently many studies report Recurrent Neural Networks as architectures of advances over classical ones. A comprehensive study with several findings about optimal settings was presented in [7]. Recurrent Neural Networks (RNN) are very efficient in text analysis, voice comparison, technical systems, medicine, etc. In [8] a type of long-short term memory model was implemented for acoustic data prediction. In [9] such structures were used for text recognition in chinese language. Medical signals of epileptic symptoms were classified by RNN as reported in [10], while in [11] these architectures helped in modeling the quality of coal fuel for power stations. In any case, the structure which uses machine learning must be trained. The second part of our study over salary prediction is devoted to best training algorithm. We have examined some variations of gradient descent approaches in our research. Such training algorithms are reported in many research, therefore we decided to examine them in our topic. In [12] was presented a study for possible training by the use of gradient descent methods. In [13] the study concentrated on devoted versions for deep learning was presented. The general model for using such algorithm in training of neural networks was defined in [14], while association rules and their relation to the training efficiency was discussed in [15]. Another aspect of efficient training is convergence analysis to the optimum while preserving the step size. The model for efficient adaptive approach was discussed in [16].

Our approach is based in RNN, which is modeled to learn from the input data about possible results from abilities of graduates related to their education, origin, experience and other social aspects. Proposed model was trained by some selected algorithms sourced in the idea of using gradient descent of the network error function in the model of weight corrections. We have performed a comparative analysis to compare efficiency of the training methods and final classification. Results are compared in charts and discussed to draw conclusions. The novelty of our approach is sourced in flexibility of our model, which is able to predict the future result from multi type input set. Additionally the structure composed for our research is easily trained and results show very good measures for many applied algorithms. We have also modified training algorithm to better fit the input values,

what is visible in increased results for such option.

## II. NEURAL NETWORK ARCHITECTURE

1) *Recurrent Neural Network*: We have done tests on different Neural Network architectures but the one that had the best accuracy was a Recurrent Neural Network using LSTM layers. Normally RNN are used for timestep data but in our case it also strongly improved accuracy. In our model we have used two activation functions:

- hiperbolic tangent - for all hidden layers,
- softmax - for the output layer.

2) *Recurrent layers*: In our model we have used Long Short-Term Memory (LSTM) layers with a forget gate. Mathematical functions used in the model are

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \quad (1)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad (2)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

$$\tilde{c}_t = \sigma_h(W_c x_t + U_c h_{t-1} + b_c) \quad (4)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \quad (5)$$

$$h_t = o_t \circ \sigma_h(c_t) \quad (6)$$

where  $c_0 = 0$  and  $h_0 = 0$  for  $x_t$  - input vector to the LSTM unit,  $f_t$  - forget gate's activation vector,  $i_t$  - input/update gate's activation vector,  $o_t$  - output gate's activation vector,  $h_t$  - hidden state vector,  $\tilde{c}_t$  - cell input activation vector,  $c_t$  - cell state vector,  $W, U, b$  - weight matrices and bias vector parameters,  $\sigma_g$  - sigmoid function,  $\sigma_c$  - hyperbolic tangent function,  $\sigma_h$  - hyperbolic tangent function.

3) *System training*: To improve our model and speed up the training process we have used an Adaptive Moment Estimation Algorithm - NAdam. This method is computationally efficient and does not require huge amount of memory so it's widely used in Machine Learning research. To improve accuracy even more and shorten training time we have used learning rate decay to firstly make bigger steps during training but after that in our approach the network is making small steps to polish the model accuracy.

Adam formula can be described as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad (7)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \quad (8)$$

where  $g$  is the current gradient value of error function for the training and  $\beta$  parameters are constant values called hyper-parameters. Values  $m_t$  and  $v_t$  are used for calculation of the correlations marked as  $\hat{m}_t$  and  $\hat{v}_t$  according to these equations:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (9)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}. \quad (10)$$

Using above calculated correlations, the final formula for changing weights in our neural network can be defined as follows:

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t, \quad (11)$$

where  $\epsilon$  is a constant small value and  $\eta$  is a learning rate (in this case we have used learning rate of 0.0003).

We need to apply NAG to Adam using these equations:

$$w_t = w_{t-1} - \eta \frac{\beta_1 m_{t-1}}{\beta_2 v_{t-1} + (1 - \beta_2) g_t^2 + \epsilon} \quad (12)$$

$$-\eta \frac{(1 - \beta_1) g_t}{\sqrt{\beta_2 v_{t-1} + (1 - \beta_2) g_t^2 + \epsilon}}$$

Now we modify Adam's update rule. The first term no longer depends on  $g_t$  so we need to change expressions for  $\hat{m}_t$  and  $w_t$ :

$$\hat{m}_t = (1 - \beta_1) g_t + \beta_{1_{t+1}} m_t \quad (13)$$

$$w_t = w_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\beta_2 v_t + \epsilon}} \quad (14)$$

## III. DATASET

The dataset for our research experiments was downloaded from kaggle.com. Initially it contained data on the education history of individual persons as well as their annual salary on the day the database was created. Some attributes, however, have been removed for processing. We removed id due to irrelevance in entering the neural network. We removed also column stating whether someone is employed or not because in our opinion this will allow the network to be used in more flexible contexts. That's why what's left is:

Gender, Secondary Education percentage- 10th Grade, Board of Secondary Grade Education, Higher Secondary Education percentage- 12th Grade, Board of Higher Secondary Education, Specialization in Higher Secondary Education, Degree Percentage, Under Graduation(Degree type)- Field of degree education, Work Experience, Employability test percentage (conducted by college), Post Graduation(MBA)- Specialization, MBA percentage, salary.

All this information was provided by graduates of Jain University Bangalore, India. According to the 4icu.org website, this university is on the 73rd place in the national classification, despite the fact that those who did not find employment and who earn less than 250k per year were in the dataset. We have established that our network will recognize this threshold. However, not only the assessments affect future employment, it is possible, among others, for two people with the same grades but different personalities, and therefore one of them will not undergo the recruitment process or negotiate a significantly lower salary.

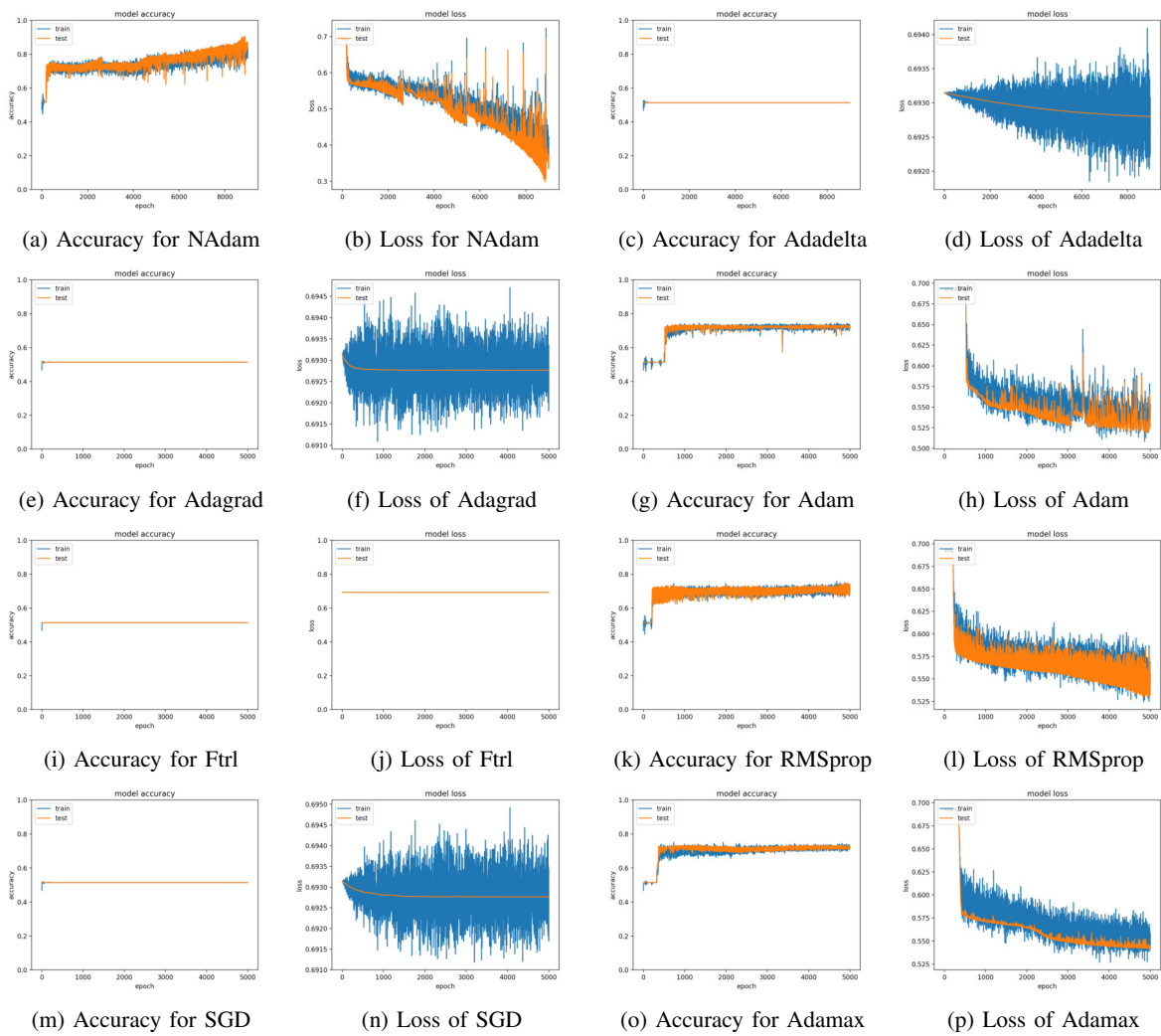


Fig. 1: Results of training using different optimization algorithms. In all charts we can see measures of accuracy and loss for the process of training.

IV. SEARCHING FOR THE BEST OPTIMIZATION ALGORITHM

Because we wanted the best accuracy of the RNN model, we have tested different learning algorithms. Some of them were better, some worse and in few cases the network did not train at all. The comparison of accuracy and loss changes during training is presented in Fig. 1. After deep analysis these are the conclusions:

- SGD - Even with double number of iterations the network categorized all values as "less then 250k".
- RMSprop - The network reached around 72.73% final accuracy.
- Adam - Network was learning but even with increased iterations maximal reached accuracy was under 75.76% and 73.14% for training data.
- Adadelta - All values were categorized as "less then 250k"
- Adagrad - All values were categorized as "less then 250k"

- Adamax - Network was learning and reached about 75.76% for test data and around 73.61% for training data. Because of that it is the second best option, however it is still a lot behind NAdam which reached around 92.42%.
- NAdam - Using NAdam gave the best results with less iterations needed.
- Ftrl - All values were categorized as "less then 250k"

Results of our initial tests on categorization are shown in Fig. 2. In each experiment we have applied classical division for training and test data 70:30.

V. CONCLUSIONS AND FUTURE WORKS

Our solution helps to understand the future and capabilities of graduates, it allows for this thanks to a well-chosen neural network training algorithm, and that's why we obtained about 92.42% final accuracy in a given dataset. As a result we can find out above all whether we should work more to improve our results through education to have opportunities in the

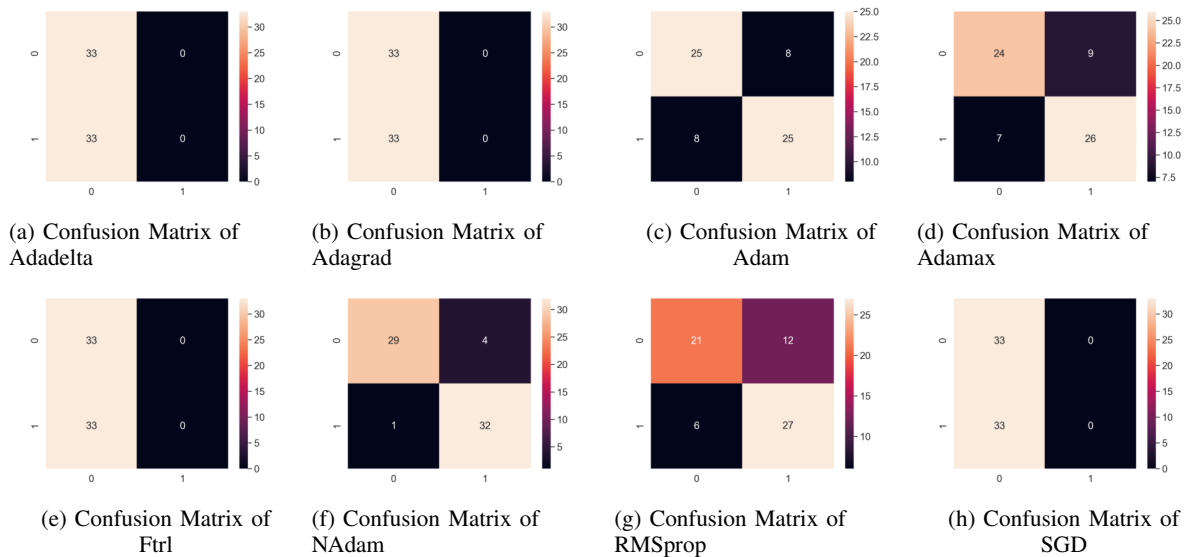


Fig. 2: Resulting confusion matrices using different optimization algorithms.

future for decent working conditions. However, it should be remembered that many factors influence the future of a given person. Despite this, behavior during your study may answer the question of what personality a person has and thus predict the fortune. In the future, we aim to further develop the project in such way that it can be determined at a young age whether a person will succeed.

In this research we have evaluated eight different training models sourced in gradient descent optimization. The results have shown that the training model of the applied RNN architecture must be also fitted to the data. Some of the methods gave very bad results and some of the methods very good, even if all of them are very similar in assumptions. Therefore in the future research we will concentrate on defining relations between input data and training procedure to achieve the best final result of classification.

#### REFERENCES

- [1] L. L. Taylor, J. N. Lahey, M. I. Beck, and J. E. Froyd, "How to do a salary equity study: With an illustrative example from higher education," *Public Personnel Management*, vol. 49, no. 1, pp. 57–82, 2020.
- [2] M. Kwiek, "Academic top earners. research productivity, prestige generation, and salary patterns in european universities," *Science and Public Policy*, vol. 45, no. 1, pp. 1–13, 2018.
- [3] M. Hernandez, D. R. Avery, S. D. Volpone, and C. R. Kaiser, "Bargaining while black: The role of race in salary negotiations." *Journal of Applied Psychology*, vol. 104, no. 4, p. 581, 2019.
- [4] E. Silva and Q. Galbraith, "Salary negotiation patterns between women and men in academic libraries," *College & research libraries*, vol. 79, no. 3, p. 324, 2018.
- [5] A. Pawha and D. Kamthania, "Quantitative analysis of historical data for prediction of job salary in india-a case study," *Journal of Statistics and Management Systems*, vol. 22, no. 2, pp. 187–198, 2019.
- [6] M. He, D. Shen, Y. Zhu, R. He, T. Wang, and Z. Zhang, "Career trajectory prediction based on cnn," in *2019 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*. IEEE, 2019, pp. 22–26.
- [7] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," *arXiv preprint arXiv:1312.6026*, 2013.
- [8] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," 2014.
- [9] J. Wang and Z. Cao, "Chinese text sentiment analysis using lstm network based on l2 and nadam," in *2017 IEEE 17th International Conference on Communication Technology (ICCT)*. IEEE, 2017, pp. 1891–1895.
- [10] A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, and D. Wunsch II, "Recurrent neural network based prediction of epileptic seizures in intra- and extracranial eeg," *Neurocomputing*, vol. 30, no. 1-4, pp. 201–218, 2000.
- [11] H. Yao, H. Vuthaluru, M. Tade, and D. Djukanovic, "Artificial neural network-based prediction of hydrogen content of coal in power station boilers," *Fuel*, vol. 84, no. 12-13, pp. 1535–1542, 2005.
- [12] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.
- [13] J. Zhang, "Gradient descent based optimization algorithms for deep learning models training," *arXiv preprint arXiv:1903.03614*, 2019.
- [14] D. Soydaner, "A comparison of optimization algorithms for deep learning," *International Journal of Pattern Recognition and Artificial Intelligence*, p. 2052013, 2020.
- [15] N. Rajesh and A. A. L. Selvakumar, "Association rules and deep learning for cryptographic algorithm in privacy preserving data mining," *Cluster Computing*, vol. 22, no. 1, pp. 119–131, 2019.
- [16] A. Barakat and P. Bianchi, "Convergence analysis of a momentum algorithm with adaptive step size for non convex optimization," *arXiv preprint arXiv:1911.07596*, 2019.