

Adaptive Regression and Classification Models with Applications in Insurance

Gints Jekabsons¹, Marina Uhanova², ^{1,2}*Riga Technical University*

Abstract – Nowadays, in the insurance industry the use of predictive modeling by means of regression and classification techniques is becoming increasingly important and popular. The success of an insurance company largely depends on the ability to perform such tasks as credibility estimation, determination of insurance premiums, estimation of probability of claim, detecting insurance fraud, managing insurance risk. This paper discusses regression and classification modeling for such types of prediction problems using the method of Adaptive Basis Function Construction.

Keywords – Classification, insurance, prediction, regression.

I. INTRODUCTION

Nowadays, in the insurance industry the use of predictive modeling by means of regression and classification techniques is becoming increasingly important and popular. The success of an insurance company largely depends on the ability to perform such tasks as credibility estimation, determination of insurance premiums, estimation of probability of claim, detecting insurance fraud, managing insurance risk etc.

It would be useful for an insurance company to have objective models that could be used to help in these and other tasks. The models can be developed based on historical data gathered over the years in the company. The intent of using a regression or classification model in these tasks is to aid the decision-making process, e.g., receive warnings if there is statistically significant evidence of potential insurance fraud.

In general, the goal of the predictive modeling is to estimate unknown (input, output) dependency (or model) from training data (consisting of a finite number of samples acquired through, e.g., experiments or simulations) with good prediction (generalization) capabilities [1], [2]. This paper discusses regression and classification modeling for such types of prediction problems using the methods of Adaptive Basis Function Construction (ABFC) [3] – [5]. The ABFC methods, in contrast to many other methods, do not have so many hyperparameters that require tedious configuration from their users, instead they automatically adapt to the particular data at hand. This property simplifies implementation and usage of ABFC in environments where most of regression and classification tasks should be automated – like in the information systems of insurance companies.

The remainder of this paper is organized as follows: Section 2 gives a brief overview of polynomial regression and classification and the subset selection approach. In

Section 3, the ABFC approach is described. Section 4 demonstrates the efficiency of the ABFC methods in a real-world insurance data analysis problem. Finally, Section 5 concludes the paper.

II. REGRESSION AND CLASSIFICATION

Predictive modeling deals with the prediction of a output variable y given the values of a vector of d -dimensional input variables $\mathbf{x} = (x_1, x_2, \dots, x_d)$. Let \mathbf{x} denote the domain of x and \mathbf{y} – the domain of y . For real-valued \mathbf{y} the problem is called regression. Otherwise, if \mathbf{y} is a finite set of unordered values, the problem is called classification. In both cases, the problem is to build a model $F(\mathbf{x})$ which maps each point in \mathbf{x} to a point in \mathbf{y} . The building of $F(\mathbf{x})$ requires a training sample of n observations (x_i, y_i) , $i = 1, 2, \dots, k$. Using the finite number of training samples, one requires to build a model $F(\mathbf{x})$ that allows predicting the output values for yet unseen input values as closely as possible.

In practical applications, most commonly a low-degree polynomial model is used. Generally, a linear regression model can be defined as a linear expansion of basis functions

$$F(\mathbf{x}) = \sum_{i=1}^k a_i f_i(\mathbf{x}), \quad (1)$$

where $\mathbf{a} = (a_1, a_2, \dots, a_k)^T$ are parameters of the model, $f_i(\mathbf{x})$, $i = 1, 2, \dots, k$ are the basis functions in the model, and k is the number of the basis functions. The model is linear in the parameters; therefore, estimation of its parameters is typically done using the Ordinary Least-Squares method minimizing the squared-error:

$$\mathbf{a} = \underset{\mathbf{a}}{\operatorname{argmin}} \sum_{i=1}^n (y_i - F(\mathbf{x}_i))^2. \quad (2)$$

On the other hand, in the classification problem, the criterion for choosing $F(\mathbf{x})$ is usually expected misclassification cost. One of the widely used tools for solving classification problems is logistic regression. Logistic regression represents log odds of y being equal to 1 as a linear model:

$$\ln(P/(1-P)) = F(\mathbf{x}) = \sum_{i=1}^k a_i f_i(\mathbf{x}), \quad (3)$$

where P is the predicted probability of y being equal to 1. P can be also represented as

$$P = 1/(1 + \exp(-F(\mathbf{x}))). \quad (4)$$

The parameters \mathbf{a} of the model are estimated by minimizing the deviance:

$$-2 \sum_{j=1}^n (y_j \ln(F(x_j)) + (1 - y_j) \ln(1 - F(x_j))) \rightarrow \min \quad (5)$$

There is no closed form solution to this minimization problem; therefore, iterative algorithms are used, e.g., Iteratively Re-weighted Least Squares [6].

III. ADAPTIVE BASIS FUNCTION CONSTRUCTION

In practical applications, the model $F(x)$ is commonly defined as a low-degree polynomial. Such a model has a small number of unknown parameters and tends to smooth out noise in the data. However, it has an important disadvantage – it may exhibit too little flexibility for modeling highly non-linear behaviors, i.e., underfit the data. Higher-degree polynomials can be used, but they may contain too many parameters and, therefore, either overfit the data or prohibit parameter estimation because the number of parameters exceeds the number of training data samples.

The most popular approach to controlling the complexity of a model is subset selection [2], [7], [8]. The goal of subset selection is from a fixed full predetermined dictionary of basis functions to find a subset that corresponds to a model of the best predictive performance.

Before performing the actual subset selection, one must first predefine the dictionary that will provide the basis functions for model generation. This is usually done by setting the maximum degree of a full polynomial and taking the set of its basis functions.

In polynomial regression, increase in the degree of a full model leads to exponential growth of the number of basis functions in the dictionary [1], [2] leading to double-exponential growth of the number of all possible subsets of basis functions. Efficient heuristic subset selection methods such as Sequential Forward Selection, SFS (also known as Stepwise Forward Selection or simply Forward Selection [2], [7]) and many others (see, e.g., [8]) considerably reduce the time. However, the time is still exponential in the degree and in the number of input variables [5].

By using the approach of subset selection one makes the assumption that the predefined fixed finite dictionary of basis functions contains a subset that is sufficient for a model to describe the target relation sufficiently well. The problem is that generally the required maximum degree is unknown beforehand and, since it will differ from one modeling task to another, it needs to be either guessed or found by additional meta-search over the whole subset selection process. In many cases (especially when the data dependencies are complex and not well understood) this means either acceptance of a possibly inadequate model or a non-trivial and long trial and error process.

In [3] – [5], an alternative approach is proposed – Adaptive Basis Function Construction. The goal of this

procedure is to overcome some of the limitations associated with the subset selection approach outlined above. ABFC is developed for building sparse polynomial regression models without restrictions on the degree of a model and enables building models in polynomial time instead of exponential time. The required basis functions are automatically constructed specifically for the dataset at hand without using a fixed predefined finite dictionary. The dictionary of basis functions in ABFC is infinite enabling generating polynomials of arbitrary complexity.

Generally, a basis function in a polynomial regression model can be defined as a product of original input variables each with an individual exponent:

$$f_i(x) = \prod_{j=1}^d x_j^{r_{ij}}, \quad (6)$$

where \mathbf{r} is a $k \times d$ matrix of non-negative integer exponents such that r_{ij} is the exponent of the j th variable in the i th basis function (note that an i th basis function can be defined as the intercept term like this: $\forall j: r_{ij} = 0$). Such a matrix with specified values for each of its elements completely defines the structure of a polynomial model with all its basis functions.

For example, if $d = 3$ and $k = 4$, then the matrix

$$\mathbf{r} = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 3 \\ 1 & 1 & 1 \end{bmatrix} \quad (7)$$

corresponds to the set

$$f = \{x_1^0 x_2^0 x_3^0, x_1^2 x_2^0 x_3^0, x_1^0 x_2^1 x_3^3, x_1^1 x_2^1 x_3^1\} \\ = \{1, x_1^2, x_2 x_3^3, x_1 x_2 x_3\} \quad (8)$$

that in turn corresponds to the model

$$F(x) = a_1 + a_2 x_1^2 + a_3 x_2 x_3^3 + a_4 x_1 x_2 x_3. \quad (9)$$

Now the problem of finding the best set of basis functions can be formally defined as finding the best matrix \mathbf{r} with the best combination of non-negative integer values of its elements:

$$\mathbf{r}^* = \operatorname{argmin}_{\mathbf{r}} J \left(\left\{ \prod_{j=1}^d x_j^{r_{ij}} \mid i = 1, 2, \dots, k \right\} \right), \quad (10)$$

where $J(\cdot)$ is an evaluation criterion that evaluates the predictive performance of the regression model corresponding to the set of basis functions.

Upper bounds of values in \mathbf{r} and value of k are not defined; therefore, it is possible to generate polynomials of arbitrary complexity, i.e., of arbitrary number of basis functions each with arbitrary exponent for each input variable.

In order to efficiently build a good regression model for a particular dataset, an efficient search mechanism is required that enables searching in an infinite space in finite time.

The search mechanism of ABFC is organized as follows (see [5] for details). The search is started from the simplest

model – the model with one basis function – the intercept term. New models are generated using so-called model refinement operators that enable adding, copying, modifying, and deleting the rows of \mathbf{r} , i.e., adding, copying, modifying, and deleting the basis functions of the model (in contrast to subset selection that restricts to just adding and deleting). Next, an efficient search strategy and a model evaluation criterion are required. In the special case of ABFC called Floating ABFC (F-ABFC), the search strategy of Sequential Floating Forward Selection [9], [10] is adapted and the corrected Akaike's Information Criterion (AICC) [11] is employed.

According to many researchers, Floating Search algorithms, including Sequential Floating Forward Selection (SFFS), are some of the most efficient deterministic search algorithms for subset selection in terms of both required computational resources and quality of results [8] – [10], [12] – [14]. Other advantages of SFFS include absence of any adjustable hyper-parameters, tendency to generate simple models, and simplicity of its implementation.

The model evaluation criterion AICC is defined as follows:

$$\text{AICC} = n \ln(\text{MSE}) + 2k + \frac{2k(k+1)}{n-k-1}, \quad (11)$$

where MSE is the Mean Squared Error of the model in the training data. AICC evaluates the predictive performance of a model as a trade-off between its accuracy in the training data and its complexity. The “best” model is then the one with the lowest AICC value. This criterion is an improvement over the classic Akaike's Information Criterion (AIC) [15] with the additional third term in (11) added as a correction term intended for small-sized datasets. For relatively small n , AICC is suited better than AIC but converges to AIC as n becomes large [11].

The termination condition of the F-ABFC search process is met, when the algorithm has generated a model that cannot be further improved using any of the refinement operators.

It should be noted that, although the state space of F-ABFC is infinite, the models of the best predictive performance are normally located in the part of the space that is relatively near to the initial state where all the models (and also their basis functions) are relatively simple, they do not yet overfit the data, and their number of basis functions is still smaller than samples in the training data. Therefore, really only a small finite fraction of the whole infinite state space is to be explored to build a “good enough” model.

A more detailed description of F-ABFC, as well as detailed empirical comparisons to subset selection methods and other popular regression modelling methods was given in our previous studies [3] – [5].

Studies in [16] – [18] demonstrated practical applications of ABFC methods for the optimum design of laser welded sandwich structures and glass fiber composite

structures. However, the methods of ABFC are really general-purpose regression modeling tools.

IV. EXPERIMENTS

This section presents the results of empirical experiments comparing the ABFC methods used with logistic regression and a number of well-known state-of-the-art classification methods. In this experiment, a real-world insurance company benchmark dataset from [19] is used. The dataset has 5822 training instances, 4000 testing instances, 85 input variables, and 1 binary output variable providing information whether or not a customer is interested in a specific insurance policy.

The methods compared are the following: 1) F-ABFC; 2) logistic regression with “full” 1st degree polynomial (FP); 3) logistic regression with sparse 1st degree polynomial built using SFS with AICC as the evaluation criterion; 4) 1, 3, and 5 Nearest Neighbors (NN); 5) Naive Bayes; 6) Multilayer Perceptron (MLP); 7) Radial Basis Function network; 8) Support Vector Machine trained using Sequential Minimal Optimization algorithm; 9) J48 classification tree [20]. Most of the compared methods are implemented in Weka tool [20] available free of charge under General Public License at <http://www.cs.waikato.ac.nz/ml/weka/>. All the methods were used with their default hyperparameters.

TABLE 1
PREDICTIVE PERFORMANCE OF THE METHODS IN TERMS OF THE NUMBER OF CORRECTLY CLASSIFIED INSTANCES (%)

Method	Correctly classified instances (%)
FP	93.98
SFS	94.00
1-NN	89.88
3-NN	93.30
5-NN	93.85
Naive Bayes	79.20
MLP	93.95
RBF	94.05
SVM	94.05
J48	93.98
F-ABFC	94.05

Table 1 presents the results of the performed experiments. As can be seen, the ABFC methods compare rather well to all the other state-of-the-art methods, in about 94% of cases giving a correct estimation of customer's interest in the insurance policy.

V. CONCLUSIONS

In this paper, the authors reviewed the ABFC approach, characterized its special case F-ABFC, and demonstrated the efficiency of the method for employment with logistic regression in a real-world classification problem from the

insurance industry, while comparing them to a number of other widely used state-of-the-art methods.

The ABFC methods, in contrast to the methods of subset selection, do not require the user to predefine the maximal degree or the dictionary of the basis functions. Instead, they automatically adapt to the particular data at hand.

Overall, the results of the performed empirical experiment are consistent with the authors' previous experiments, and it can be concluded that the adaptive model building methods of ABFC are able to give results that are comparable to results of other state-of-the-art methods.

While the performance of most of the used methods is similar, it should be noted that the ABFC methods, in contrast to many other methods, do not have so many hyperparameters that require tedious configuration from their users. Instead, they automatically adapt to the particular data at hand. This property simplifies implementation and usage of ABFC in environments where most of regression and classification tasks should be automated, e.g., in the information systems of insurance companies.

The future research will focus on more thorough studies and experiments regarding ABFC efficiency in the insurance industry.

ACKNOWLEDGEMENTS

The research is partly funded by the ERDF (ERAF) project (No. 2011/0008/2DP/2.1.1.1.0/10/APIA/VIAA/018 "Development of Insurance Distributed Software Based on Intelligent Agents, Modeling, and Web Technologies")

REFERENCES

- [1] V. Cherkassky, and F. M. Mulier, *Learning from Data: Concepts, Theory, and Methods*, 2nd ed., Wiley-IEEE Press, 2007. <http://dx.doi.org/10.1002/9780470140529>
- [2] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2003.
- [3] G. Jekabsons, "Ensembling Adaptively Constructed Polynomial Regression Models," *International Journal of Intelligent Systems and Technologies (IJIST)*, vol. 3, no. 2, pp. 56-61, 2008.
- [4] G. Jekabsons and J. Lavendels, "Polynomial Regression Modelling using Adaptive Construction of Basis Functions," *Proceedings of IADIS International Conference, Applied Computing 2008, Algarve, Portugal*, pp. 269-276, 2008.
- [5] G. Jekabsons, "Adaptive Basis Function Construction: An Approach for Adaptive Building of Sparse Polynomial Regression Models," in *Machine Learning*, Y. Zhang, Ed. In-Tech, 2010, pp. 127-156.
- [6] P. Komarek and A. Moore, "Making Logistic Regression A Core Data Mining Tool With TR-IRLS," *Proceedings of the 5th International Conference on Data Mining Machine Learning*, 2005.
- [7] J. O. Rawlings, *Applied Regression Analysis: A Research Tool*, 2nd ed., Pacific Grove, CA: Wadsworth & Brooks/Cole, 1998. <http://dx.doi.org/10.1007/b98890>
- [8] J. Reunanen, "Search Strategies," *Feature Extraction: Foundations and Applications*, I. Guyon, S. Gunn, M. Nikravesh, L. A. Zadeh, Eds. Springer, pp. 119-137, 2006. http://dx.doi.org/10.1007/978-3-540-35488-8_5
- [9] P. Pudil, F. J. Ferri, J. Novovicova, and J. Kittler, "Floating search methods for feature selection with nonmonotonic criterion functions," *Proceedings of the International Conference on Pattern Recognition*, vol. 2, Los Alamitos, CA: IEEE, pp. 279-283, 1994.
- [10] P. Pudil, J. Novovicova, and J. Kittler, "Floating Search Methods in Feature Selection," *Pattern Recognition Letters*, vol. 15, pp. 1119-1125, 1994. [http://dx.doi.org/10.1016/0167-8655\(94\)90127-9](http://dx.doi.org/10.1016/0167-8655(94)90127-9)
- [11] C. M. Hurvich and C-L. Tsai, "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76, pp. 297-307, 1989. <http://dx.doi.org/10.1093/biomet/76.2.297>
- [12] A. K. Jain, D. P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, 2000. <http://dx.doi.org/10.1109/34.824819>
- [13] D. Zongker and A. Jain, "Algorithms for feature selection: an evaluation," *Pattern Recognition*, vol. 2, pp. 18-22, 1996.
- [14] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognition*, vol. 33, no. 1, pp. 25-41, 2000. [http://dx.doi.org/10.1016/S0031-3203\(99\)00041-2](http://dx.doi.org/10.1016/S0031-3203(99)00041-2)
- [15] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, AC-19, pp. 716-723, 1974. <http://dx.doi.org/10.1109/TAC.1974.1100705>
- [16] K. Kalnins, E. Eglitis, G. Jekabsons, and R. Rikards, "Metamodels for Optimum Design of Laser Welded Sandwich Structures," *Scientific Proceedings of International Conference on Welded Structures, Design, Fabrication, and Economy 2008*, Miskolc, Hungary, pp. 119-126, 2008.
- [17] K. Kalnins, O. Ozolins, and G. Jekabsons, "Metamodels in Design of GFRP Composite Stiffened Deck Structure," *Proceedings of 7th ASMO-UK/ISSMO International Conference on Engineering Design Optimization*, Association for Structural and Multidisciplinary Optimization in the UK, Bath, UK, 11 p., 2008.
- [18] K. Kalnins, G. Jekabsons, K. Zudrags, and R. Beitlers, "Metamodels in optimisation of plywood sandwich panels," *Shell Structures: Theory and Applications*, W. Pietraszkiewicz, C. Szymczak, Eds. CRC Press, pp. 291-294, 2009.
- [19] P. van der Putten and M. van Someren, Eds. "CoIL Challenge 2000: The Insurance Company Case," *Sentient Machine Research*, Leiden Institute of Advanced Computer Science, Amsterdam, The Netherlands, Tech. Rep. 2000-09, 2000.
- [20] I. H. Witten, E. Frank, *Data mining: practical machine learning tools and techniques with Java implementations*, 2nd ed., SF: Morgan Kaufmann, 2005.

Gints Jekabsons graduated from Riga Technical University, receiving the Bachelor's Degree in 2003 and the Master's Degree in 2005. He received an Award of the Latvian Fund of Education for the best Bachelor's Thesis in 2003 and the best Master's Thesis in 2005. In 2009, he obtained the Doctoral Degree at Riga Technical University. The topic of the Doctoral Thesis was "Heuristic Methods in Multidimensional Regression Model Building". The Doctoral Thesis won the Verner fon Siemens Excellence Award in 2009. Since 2005, he has worked as a Lecturer, Programmer, and Researcher. His current employment is a Leading Researcher at the Institute of Applied Computer Systems, Riga Technical University. His research interests include machine learning, statistics, and location-based services. E-mail: gints.jekabsons@rtu.lv

Marina Uhanova graduated from Riga Aviation University, receiving the Bachelor's Degree in 1995 and Master's Degree in 1996. In 2007, she received the Doctoral Degree in System Analysis, Modeling and Development from Riga Technical University. Since 2008, she has worked as an Assistant and Lecturer. Her current employment is an Assistant Professor at the Chair of Informatics and Programming, Riga Technical University. Her research interests include distributed application development and their applications in insurance. E-mail: marina.uhanova@rtu.lv