A Classroom Observation Model Fitted to Stochastic and Probabilistic Decision Systems

Marios Poulos, Vassilios S. Belesiotis, Nikolaos Alexandris

A Classroom Observation Model Fitted to Stochastic and Probabilistic Decision Systems

Marios Poulos[1], Vassilios S. Belesiotis[2], Nikolaos Alexandris[3,1]

[1] Ionian University, Department of Archives and Library Science, Ioannou Theotoki 72, 49100, Corfu Greece
[2,3] University of Piraeus – Department of Informatics
mpoulos@ionio.gr, vbel@unipi.gr, alexandr@unipi.gr

Abstract. This paper focuses on solving the problems of preparing and normalizing data that are captured from a classroom observation, and are linked with significant relevant properties. We adapt these data using a Bayesian model that creates normalization conditions to a well fitted artificial neural network. We separate the method in two stages: first implementing the data variable in a functional multi-factorial normalization analysis using a normalizing constant and then using constructed vectors containing normalization values in the learning and testing stages of the selected learning vector quantifier neural network.

Keywords: Probabilistic Decision Systems, Normalization Data, Education, Formative Assessment in the Classroom, Classroom Observation.

1 Introduction

1.1 Background and Motivation

The preparation stage of data processing is the most significant procedure in stochastic and probabilistic decision systems [1], an important subclass of which is neural networks. Modeling has an enormous impact on the success of a wide range of such complex data analyses as data mining and feature-extraction procedures [1, 2], mainly because the quality of the input data in neural network models strongly influences their data analysis's results [1, 3] and the efficiency of their performance, as wrongly prepared data is likely to make the data analysis problematic [1]. The appropriate pre-processing of input data is therefore a significant procedure [1, 4]. A study of the procedure for grouping collected data [1] that addresses the selection of

[1] N. Alexandris. Visiting Professor at the University of Canberra, Faculty of Information Sciences and Engineering, is Professor of University of Piraeus in the Department of Informatics.

the data's variables with genetic algorithms and those algorithms' linkage to significant relevant properties would therefore be valuable.

1.2 Problems with the Preparation Data Stage

Grouping these categories with the specific multi-scale and normalization practices in the non-structural data's category may, however, be overly simplistic [1], as doing so cannot provide solutions for significant multi-factorial unconstructed data that are linked with significant relevant properties. A characteristic application of this method is the domain of classroom observation [5], [6], because data collected from this process fall into the aforementioned category.

Existing solutions have focused on the consolidation phase, during which data management systems consist of a continuous process. The capture procedure stores unconstructed data in a well-constructed database [7], but their basic problem is in the pre-processing training setting, as the number of possible interaction factors grows exponentially with the inclusion of each additional main effect in the logistic regression model. Logistic regression's ability to deal with interactions involving many factors is therefore limited, as having too many independent variables in relation to the number of observed outcome events is problematic [8], [9].

1.3 Aims and Scope

This paper attempts to solve the problems of preparing and normalizing data linked with significant relevant properties and captured from many sources by adapting these factors to a first phase using a Bayesian model that creates normalization conditions with an equitable distribution between the significant relevant properties of such types of data. It also aims to adapt this normalization to a well-fitted neural network so it can accept pre-processing data for training and testing procedures. It does this using two separate procedures, first implementing the data variable in a functional multi-factorial normalization analysis using a normalizing constant and then using constructed vectors containing normalization values in the learning and testing stages of the selected learning vector quantifier neural network.

For application purposes of this model we focus on the educational domain and more specifically on classroom observations in order to construct a decision system able to learn to evaluate the quality of a teaching process given as input a set of observations describing the events that took place during this process. The required responses of the system are boolean (yes or no) and form a statistical model which follows the t-test. The proposed method will focus on the normalization of the input data, enabling the effective training of a neural network in order to provide evaluation results for any given set of data. The evaluation of a teaching process can be any value from the fuzzy set: excellent, good, adequate, poor.

This paper is divided in four sections. Section 2 describes the integrated data preparation method as well as the architecture of the neural network with the original classroom observation model. In Section 3, the fitted model is presented, adapting the

model to the classroom observation data. Finally, Section 4 concludes and explains our plans for future work.

2 Proposed Method

Figure 1 outlines the schedule of this approach. Following our method, captured classroom observation data are normalized using a Bayesian model and adapted to a well-fitted neural network.

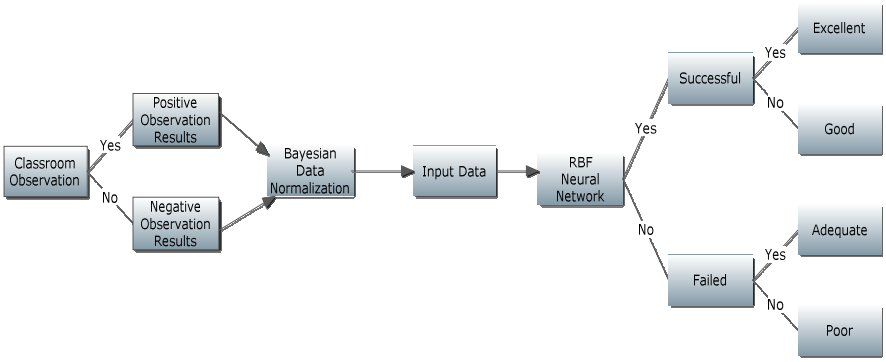

Fig. 1. The schedule of the proposed method is analyzed

2.1 Education Model

The original model is based on Systematic Observation [8], which presents a process for categorizing classroom observations regarding events describing the teacher's and pupils' behavior in a well constructed form. The pioneer of this method Flanders introduced a system [9] which has formed the basis for a number of studies in this field [8]. He defines teaching behavior as 'acts by the teacher which occur in the context of classroom interaction' [9]. Flanders devised a system which is divided into 3 broad areas: teacher talk, pupil talk and silence. This classification is overviewed in Table 1. However, in order to apply our method we consider that the expected value for the number of occurrences of each characteristic is equal to S/n, where S is the total number of observations and n is the number of characteristics. Furthermore, we consider a confidence probability of a=0.025, ranging the area between $x^2_{0,975,v}$ and $x^2_{0,025,v}$ with degree of freedom v=10-1=9.

Table 1: Flanders' Interaction Analysis categories [9: 34], where T: Teacher, R: Response, I: Initiation, S is the total number of observations and S_i (i=1..9) is the number of observations for each characteristic.

Parameters	Characteristics	Groups	Expected Result Total=S	Observation Result Total=S A=0.025
1	**Teacher, accepts feeling**. Accepts and clarifies an attitude or the feeling tone of a pupil in a non-threatening manner. Feelings may be positive or negative. Predicting and recalling feelings are included.	T-R	S/10	S1/S
2	**Teacher, Praises or encourages**. Praises or encourages pupil action or behaviour. Jokes that release tension, but not at the expense of another individual: nodding head, or saying 'Um hm?' or 'Go on' are included.	T-R	S/10	S2/S
3	**Teacher, accepts or uses ideas of pupils**. Clarifying, building or developing ideas suggested by a pupil. Teacher extensions of pupil ideas are included but as the teacher brings more of his own ideas into play, shift to category five.	T-R	S/10	S3/S
4	**Teacher, Asks questions**. Asking a question about content or procedure, based on teacher ideas, with the intent that a pupil will answer.	T-I	S/10	S4/S
5	**Teacher, Lecturing**. Giving facts or opinions about content or procedures: expressing his own ideas, giving his own explanation or citing an authority other than a pupil.	T-I	S/10	S5/S
6	**Teacher, Giving directions**. Directions, commands or orders to which a pupil is expected to comply.	T-I	S/10	S6/S
7	**Teacher, Criticizing or justifying authority**. Statements intended to change pupil behavior from non-acceptable to acceptable pattern; bawling someone out; stating why the teacher is doing what he is doing; extreme self-defense.	T-I	S/10	S7/S
8	**Pupil talk – response**. Talk by pupils in response to teacher. Teacher initiates the contact or solicits pupil statement or structures the situation. Freedom to express own ideas is limited.	P-R	S/10	S8/S
9	**Pupil talk – initiation**. Talk by pupils which they initiate. Expressing own ideas; initiating a new topic; freedom to develop opinions and a line of thought, like asking thoughtful questions: going beyond the existing structure.	P-I	S/10	S9/S
10	**Silence or confusion**. Pauses, short periods of silence and periods of confusion in which communication cannot be understood by the observer.	S		

2.2 Integrated Data Preparation Method

According to previous study [1], the proposed technique to formulate related features of non-structural data is based on extracting related features from text data or documents by semantic analysis and formulates an event-specific summary.

But this technique doesn't support our cases which are based on % related results. In this case, the solution of this conversion could be found in the Bayes's theorem according to the posterior probability measure is proportional to the product of the prior probability measure and the likelihood function. Proportional to implies that one must multiply or divide by a normalizing constant to assign measure 1 to the whole space, i.e., to get a probability measure. In a simple discrete case we have

$$P(H_0|D) = \frac{P(D|H_0)P(H_0)}{P(D)} \tag{1}$$

where $P(H_0)$ is the prior probability that the hypothesis is true; $P(D|H_0)$ is the conditional probability of the data given that the hypothesis is true, but given that the data are known it is the likelihood of the hypothesis (or its parameters) given the data; $P(H_0|D)$ is the posterior probability that the hypothesis is true given the data. P(D) should be the probability of producing the data, but on its own is difficult to calculate, so an alternative way to describe this relationship is as one of proportionality:

$$P(H_0 \mid D)P(D \mid H_0)P(H_0). \tag{2}$$

Since P(H|D) is a probability, the sum over all possible (mutually exclusive) hypotheses should be 1, leading to the conclusion that

$$P(H_0|D) = \frac{P(D|H_0)P(H_0)}{\sum_i P(D|H_i)P(H_i)}. \tag{3}$$

In this case, the reciprocal of the value

$$P(D) = \sum_i P(D|H_i)P(H_i) \tag{4}$$

In our case according to Table 1 the prior probability is given by $P(H_0)$ =S1/S. The first conditional probability is given by $P(D|H_0) = P(H_0)/\alpha*100$.

$$k(j)=1/P(D) \tag{5}$$

is the normalizing constant [10] where j=1..9 is number of the parameters. Vector k(j) is of size 1xj.

2.3 Neural network Architecture

We adopt as the ideal neural network to accept an appropriate fitted Bayesian vector k(j), since the main applications of RBF have been shown to be in pattern with a Bayesian classifier [11]. For this purpose we select an RBF neural network classifier from a variety of available neural networks architectures, for the following reasons:

The RBF network was preferred to a Multilayer Perceptron (MLP), because it can solve a given problem using fewer neurons and in sorter time, yet with the same success.

The RBF network has the ability to classify incoming vectors into classes that are not linearly separable because it uses static Gaussian function as the nonlinearity for the hidden layer processing elements [12]. The Gaussian function responds only to a small region of the input space where the Gaussian is centered. Successful application of this classifier relies on finding suitable centers for the Gaussian functions. An unsupervised approach usually produces better results than supervised learning.

Using n classes according to the referred problem we design the node characteristics and the RBF network topology.

An RBF neural network can be considered as a special three-layered network. The input nodes pass the input values to the internal nodes that formulate the hidden layer. The nonlinear responses of the hidden nodes are weighted in order to calculate the final outputs of network in the third (output) layer [13].

A typical hidden node in an RBF network is characterized by its center, which is a vector with dimension equal to the number of inputs to the node. The activity νl(x) of the lth node is the Euclidean norm of the difference between the input vector k(j) and the node center and is given by:

$$\nu_i(x) = \left\| x - \hat{x}_i \right\| \tag{1}$$

The output function of the node is a radially symmetric function. A typical choice, which is also used in this work, is the Gaussian function [14]:

$$f(\nu) = \exp\left(-\frac{\nu^2}{\sigma^2} \right) \tag{2}$$

where σ is the width of the node.

3 Fitted Model

According to Sections 2.1, 2.2 and 2.3 we constructed the vector of the normalization data k(j). For the training procedure we constructed a satisfied number n vectors k(j), where n-h (h<n h E N) is the number of vectors used in the training procedure. It must be noted that the number of the vectors n is distributed in the four target classes of the RBF neural network. These classes correspond to a decision Liker scale

(excellent, good, satisfied, fair). For implementation purposes we used 4 neurons and 2 layers. In the testing procedure we used the sigmoid function and h vectors k(j).

For the statistical evaluation of this model we used t-test criterion in order to evaluate the significant probability by considering as hull hypothesis that the system yields a random results aiming to the significant rejection.

4 Conclusion-Futures

In this paper we focus on a method solving problems related to preparing and normalizing data linked with significant relevant properties, which have been captured from classroom observations by adapting these factors to a first phase using a Bayesian model that creates normalization conditions with an equitable distribution between the significantly relevant properties of such types of data. Moreover, we described how the normalization can aid the training of a well-fitted neural network.

In the future, we would like to perform an extensive statistical evaluation of our model with real classroom observation data obtained through experimental questioners. This process can be supported by the software environment proposed in [14], which provides a system aiding the recording of observed classroom events. Finally, we would like to extend [14] by enhancing it with decision making capabilities in order to help the teacher identify problems and provide formative assessment.

References

1. Yu, L., Wang, S., Lai, K. K.: An integrated data preparation scheme for neural network data analysis, Knowledge and Data Engineering, IEEE Transactions 18(2), 217--230 (2006)
2. Hu, X.: DB-H Reduction: A Data Preprocessing Algorithm for Data Mining Applications, Applied Math. Letters 16, 889--895 (2003)
3. Sattler, K.U., Schallehn, E.: A Data Preparation Framework Based on a Multidatabase Language, Proc. Int'l Symp. Database Eng. & Applications, 219--228 (2001)
4. Pastor-Satorras, R., Vespignani, A.: Epidemics and immunization in scale-free networks. In: Bornholdt S., Schuster H. G. (eds.), Handbook (2002)
5. IATEFL, www.ttedsig.iatefl.org/resources/articles/6.doc (attached 29-3-2010)
6. Flanders, N.A.: Analyzing Teaching Behaviour. Addison Wesley (1970)
7. Tsumoto, S.: Mining diagnostic rules from clinical databases using rough sets and medical diagnostic model. Information Sciences: An International Journal 162(2), 65--80 (2004)
8. Ritchie, M.D., Motsinger, A.A., Bush W.S., Coffey C.S., Moore J. H.: Genetic Programming neural networks: A powerful bioinformatics tool for human genetics, Applied Soft Computing 7, 471--479 (2007)
9. Moore, J.H., Williams, S.M.: New strategies for identifying gene–gene interactions in hypertension, Ann. Med. 34, 88--95 (2002)
10. Feller, W.: An Introduction to Probability Theory and its Applications (volume I). John Wiley & Sons. ISBN 0-471-25708-7 (1968)
11. Traven, H.G.C.: A neural-network approach to statistical pattern classification by semiparametric estimation of a probability density functions. IEEE Trans. Neural Networks 2, 366--377 (1991)

12. Anita, S.Y., et al.: Transvaginal ultrasound measurement ofendometrial thickness as a biomarker for estrogen exposure. Cancer Epidemiology Biomarkers 13(9), 1459--65 (1968)
13. Sarimveis, H., Doganis, P., Alexandridis, A. et al.: A classification technique based on radial basis function neural networks. Advances in Engineering Soft2ware 37(4), 218--221 (2006)
14. Belesiotis, V.: A software system for the didactic utilization of classroom observation for Informatic courses (in Greek): 5th Panhellenic Conference on Didactic on Informatics, pp. 259-256 (2010)