Preventing the Impact of Marital Dissolutions in Children by Regression Techniques

Nuria Rico[1], Alberto Guillén[2], Carlos Tovar[3], and José F. Guillén[3]

[1] Department of Statistics and Operational Research
nrico@ugr.es
[2] Department of Computer Architecture and Computer Technology,
Universidad de Granada, Spain
aguillen@atc.ugr.es
[3] Department of Preventive Medicine
carlos.tovar@juntadeandalucia.es, fguillen@ugr.es

Abstract. The process of marital dissolution is a crisis that affects both the couple and their offspring. Many studies have shown how children involved in a marital dissolution could present less adaptation abilities as well as less healthy live habits. The longer the process, the more serious become these problems. Therefore, to be able to take preventive actions would be quite useful towards minimizing the dissolution process impact. This paper aims at supporting the decision of doctors when deciding about a possible treatment to children involved in a dissolution process studying the extension of time that the dissolution procces spend.

Classical statistical techniques as well as latest machine learning algorithms will be applied in order to predict how long the dissolution might take and which parameters could be the most significant. The information used in this study comes from the Spanish government monitorization of the dissolutions during the last years.

1 Introduction

Diseases and health problems are not randomly distributed in society. Scientific research on public health field widely shows that social and demographic factors have enormous influence on population and people health [6]. A large number of studies prove that health inequalities derived from origin and social stratum explain mortality and morbidity excess more than the known specific risk factors for the most of diseases [7], [8].

Several medical and sociological researchs show that marital dissolution are a powerful psycho-social risk factor which can make high deteriorations on people involved health, both adult and children, as in physical as mental level (see [9], [10], [11], [12], [13], [14], [15]). The studies with vital trajectories show that the negative effects due to marital crisis and transitions on health are accumulated and extended for the rest of life [16].

There exists a lot of scientific evidences about the break of a couple is associated with a higher risk on the children of the couple showing alterations,

M.M. Cruz-Cunha et al. (Eds.): CENTERIS 2011, Part III, CCIS 221, pp. 264–271, 2011.

disorders and loss in health (see [17]). Longitudinal studies demonstrate that worse initial situations, during childhood and youth, more severely limite the opportunities to achieve optimum levels on health. The adverse psycho-social circumstances, emphasizing the family breakdown, do exert a direct negative influence on the socio-economic situation, which is on the base of the explanation for every health inequality models.

Time duration of the legal process for marital dissolution behave as a worst forecast factor in terms of the magnitude, intensity and duration of the negative effects observed in the children health and welfare.

A way to prevent the negative effects on health is to reduce the duration of the process, understanding that higher time duration, higher the adverse effect is. We have studied how is time duration of the process related with other variables that the National Statistics Institute of Spain collect, for public statistic aims, with the objective to find a function that support medical decisions.

The results showed in this paper allow to forecast rather the duration time of the dissolution process will be either less or longer than six months. In section 2 the objectives of the research are established. Following, in section 3 the material and methods used are described and in section 4 the experimental results are exposed and briefly analyzed. Finally, in section 5 conclusions of the research are summarized.

2 Objectives

This paper is focused on forecasting the time duration of the marital dissolution processes. For this aim we will use two prediction models under two different points of view: the first one is made by discriminant analysis, as an appropriate statistical technique, and the other is carried out with RBF Neuronal Network, as a recently developed and powerful computational technique.

The output of the system could be incorporated in a decision support system in order to identify the possibility of going through a large dissolution process. If this is possible, preventive procedures could be taken in order to minimize the damage the people involved in the dissolution might suffer.

3 Material and Methods

For forecasting the dissolution process time duration, two methodologies have been carried out: the fist one applying discriminant analysis and the other one applying RBF Neuronal Network. Both have been developed using the data base provided by the National Statistic Institute of Spain that we describe in section below. Subsections following briefly describe the prediction methods and software required.

3.1 Data Base

To be agreed with the objectives of the research, we have developed forecast mechanism for the time duration of the marital dissolution process, based on certain information related to family socio-demographic characteristics. The database set used in the estimation and validation of the forecasting models is the microdata file about all the dissolution sentences passed on Spain during 2008. This data set, provided by the Judge Statistics Unit on the National Statistics Institute of Spain, collect all computed sentences by the General Council Judiciary between, January-1th and December-31th on 2008.

All the 81855 cases are described in the database set by 30 variables with a high level of enforceability. Valid data are between 95% and 100%. In the database homosexual marital dissolutions are excluded. For each pass, the variables computed are: Province and Autonomous Region; Birth Day, Month and Year for each spouse, Day, Month and Year for the marriage holding, of the lawsuit bring and of the sentence pass; the Plaintiff Spouse; Nationality and Previous Marital Status for each married couple; Previous Marital Separation; Number of under 18s children; Who pays Compensatory Damages; Who pays Maintenance; Who has under 18s children custody; and Judgement Declaration.

This database provides us any amount of socio-demographic information for families with a marital dissolution occurred in 2008. From these data information we have generated some new variables, removing the reference year effect, and we have used them for forecasting the time duration of the process. These variables are: Husband and Wife ages at marriage and the time between the marriage and the dissolution lawsuit. Therefore, we have taken into account the contentiousness generating a new variable. It has been possible because Judgement Declaration has the necessary information. Finally, we have removed from the independent variables data set the dates of the lawsuit bring and sentence pass. The first one is the dependent variable which we want to predict, and the later one gives, since we know that all sentences are passed in 2008, information about the process duration. This information might not be take because it is not available usually. After the transformations on the database, we have the following set of prediction variables:

Variable	Socio-demographic information related
V1	Province
V2	Number of under 18s children
V3	Husband Nationality
V4	Husband Previous Marital Status
V5	Wife Nationality
V6	Wife Previous Marital Status
V7	Judgement Declaration
V8	Plaintiff Spouse
V9	Previously Marital Separation
V10	Autonomous Region
V11	Time between Marriage and Dissolution lawsuit
V12	Husband age at marriage
V13	Wife age at marriage
V14	Contentiousness

The data set has been divided into two subsets: the named training data set and the named test data set. The purpose of this division is to make the prediction model using the first one and to check the models efficiency on the second data set. For this aim, we have randomly divide the 81855 cases into 46895 cases (57.29%) for training set and 34960 cases (42.71%) for test set.

3.2 Discriminant Analysis

Discriminant Analysis is a statistical technique developed by Fisher [1]. It is a multivariate statistic technique focused on verify if significant differences between groups exist, explain the differences and forecast the most probably group for new cases (see [2]). There must be a variables set, observed in these groups to explain, if the differences exist, how they are and which way they occur, giving us a systematic classification procedure for new observations. This Analysis looks for discriminant functions, calculated in a similar way that multiple linear regression equations. It consists in getting, from the independent variable set, some lineal functions $D_k = u_1 x_{1k} + u_2 x_{2k} + \ldots + u_d x_{dk}$ with $k = 1, \ldots, n$ with the power to classify another observations.

In the application of discriminant analysis, it has been used SPSS software, in the 15th version. This program has the advantages of intuition and easy to work with. The most important disadvantage consists in the fine depuration that it must be done on the data set before running the program. Furthermore, manipulate and insert modifications on the process are not simple.

3.3 RBFNN Description

A Radial Basis Function Neural Network (RBFNN) is a especial kind of Artificial Neural Networks which has an unique hidden layer of processing units, and each one of those, computes a Radial Basis Function. This kind of networks are very popular since they are able to approximate any function [21]. Given a designed RBFNN, the output is computed as a weighted sum of the outputs of the neurons:

$$\mathcal{F}\left(\boldsymbol{x}_k; C, R, \Omega\right) = \sum_{j=1}^{m} \phi(\boldsymbol{x}_k; \boldsymbol{c}_j, r_j) \cdot \Omega_j \tag{1}$$

where $C = \{\boldsymbol{c}_1, ..., \boldsymbol{c}_m\}$ is the set of RBF centers, $R = \{r_1, ..., r_m\}$ is the set of values for each RBF radius, $\Omega = \{\Omega_1, ..., \Omega_m\}$ is the set of weights and $\phi(\boldsymbol{x}_k; \boldsymbol{c}_j, r_j)$ represents an RBF. For regression and classification problems, due to its properties and good behaviour, the most commonly used RBF is the Gaussian function:

$$\phi(\boldsymbol{x}; \boldsymbol{c}_i, r_i) = \exp\left(\frac{-||\boldsymbol{c}_i - \boldsymbol{x}||^2}{r_i^2}\right) \;, i = 1, ..., m \tag{2}$$

Design of an RBFNN. As described above, there is a set of parameters that have to be initialized in order to obtain an RBFNN, these are: number of RBFs, their position in the solution space, their widths and the weights for the output layer. Although there are several methodologies to design RBFNN [18,20], the classical procedure consists of:

1. Initialize RBF centers $\boldsymbol{c}_j$
2. Initialize the radius r_j for each RBF
3. Calculate the optimum value for the weights Ω_j
4. Apply a local search algorithm

Once the centers are initialize, the widths are usually computed based on their position controlling the overlapping degree. Afterwards, the output weights can be calculated optimally by solving a linear equation system.

The first step of initialization is crucial in order to obtain good approximation results due to the large number of local minima. In order to overcome this problem, specific clustering algorithms have been developed [24,23,19]. In this paper we have selected the approach proposed in [19] due to its good results and possibility of obtain with the same algorithm the widths for the RBFs.

This algorithm defines a distortion function that has to be minimized in order to make an adequate placement of the centers:

$$\delta = \sum_{k=1}^{n} \sum_{i=1}^{m} D_{ik}^2 a_{ik}^l |Y_k^p| \tag{3}$$

where D_{ik} represents the Euclidean distance from a center c_i to an input vector $\boldsymbol{x}_k$, a_{ik} is the activation value that determines how important the input vector $\boldsymbol{x}_k$ is for the center $\boldsymbol{c}_i$, l is a parameter to control the degree of overlapping between the neurons, Y_k is the preprocessed output of the input vector $\boldsymbol{x}_k$, and p allows the influence of the output when initializing the centers to increase or decrease.

4 Experimental Results

This section will compare the regression techniques in order to see which one performs better in general and also analysing the specificity. The marital dissolutions were classified in two groups: less than 6 months and more than 6 months. The results obtained by the Discriminant Analysis are shown in Table 1 and the ones provided by the RBFNN are shown in Table 2. The continuous output of the RBFNN is shown in Fig. 1 where it is possible to see how is able to approximate reasonable well the marital dissolutions between the range of 100 to 300 days although the others, specially the largest and the smallest, is not able to perform an accurate approximation.

Both methods perform well, making a correct classification in most of the cases although the approximation of the number of days is not too accurate. Regarding the accuracy of the classification, it is possible to see how it more difficult to detect if a dissolution will last over 6 months. Nonetheless, both models obtain accuracies over the 60%, which is an acceptable result.

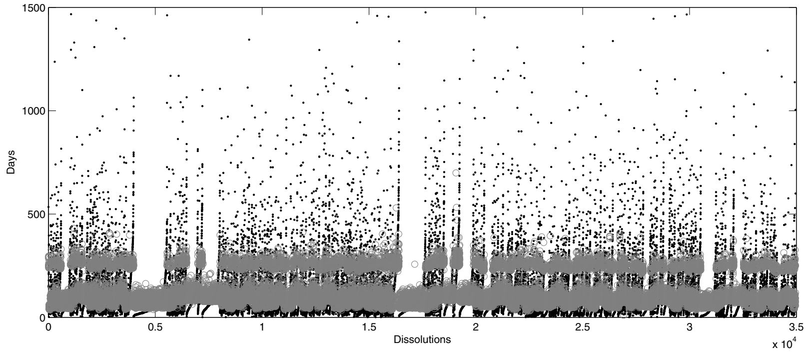

Fig. 1. Approximation of the RBFNN over the test data set. Black dots are the real output and grey circles are the output of the network.

Table 1. Classification Results for the Discriminant Analysis

Training	Total	Less than 6 months	More than 6 months
Correct classification	82,7%	87,96%	64.48%
Wrong classification	17,3%	12,14%	35.52%
Test	Total	Less than 6 months	More than 6 months
Correct classification	83%	87.94%	65.7%
Wrong classification	17%	12.06%	34.3%

Table 2. Classification Results for the RBFNN using X neurons

Training	Total	Less than 6 months	More than 6 months
Correct classification	82.65%	88.1%	64.4%
Wrong classification	17.35%	11.9%	35.6%
Test	Total	Less than 6 months	More than 6 months
Correct classification	82.9%	88.05%	65.6%
Wrong classification	17.1%	11.95%	34.4%

5 Conclusions

Marital dissolutions are a traumatic problem for all the people involved in them that directly affects their health. As the longer this process becomes, the more severe the damages are, it is useful to have a decision support system that, given a set of variables, is able to determine if the dissolution will last for a long time or not so preventive methods can be applied. This paper has presented two mathematical models that are able to asses about the duration of the dissolutions. Both approaches, the statistical and the machine learning, have performed similarly showing no significant differences regarding the accuracy of the results.

Acknowledgments. This research has been supported by the projects by the Spanish CICYT Project TIN2007-60587 and TEC2008-04920 and Junta Andalucia Projects P08-TIC-03674 and P08-TIC03928 and PYR-2010-17 of CEI BioTIC GENIL (CEB09-0010) of the MICINN.

References

1. Fisher, R.A.: The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics 7(2), 179–188 (1936), `http://hdl.handle.net/2440/15227`, doi:10.1111/j.1469-1809.1936.tb02137.x (retrieved 2009-05-09)
2. McLachlan, G.J.: Discriminant Analysis and Statistical Pattern Recognition. Wiley Interscience, Hoboken (2004) ISBN 0471691151. MR1190469
3. Anda, R.F., Butchart, A., Felitti, V.J., Brown, D.W.: Building a Framework for Global Surveillance of the Public Health Implications of Adverse Childhood Experiences. American Journal of Preventive Medicine 39(1), 93–98 (2010)
4. Chartiera, M.J., Walkerb, J.R., Naimark, B.: Separate and cumulative effects of adverse childhood experiences in predicting adult health and health care utilization. Child Abuse & Neglect 34, 454–464 (2010), doi:10.1016/j.chiabu.2009.09.020
5. Due, P., Krølner, R., Rasmussen, M., Andersen, A., Damsgaard, M.T., Graham, H., Holstein, B.E.: Pathways and mechanisms in adolescence contribute to adult health inequalities. Scandinavian Journal of Public Health 39, 62 (2011), doi:10.1177/1403494810395989
6. Benach, J.: Los determinantes demográficos y sociales de la salud. Gaceta Sanitaria 18(supl. 1), 7 (2004)
7. Linj, B.G., Phelan, J.: Social conditions as fundamental causes of diseases. Journal of Health and Social Bahavior 24, 80–94 (1995)
8. Borrell, C., Artazcoz, L.: Las políticas para disminuir las desigualdades en salud. Gaceta Sanitaria 22(5), 465–473 (2008)
9. Ebrahim, S., Wannamethee, G., McCallum, A., Walker, M., Shaper, A.G.: Marital status, change in marital status, and mortality in middle-aged british men. American Journal of Epidemiology 142, 834–842 (1995)
10. Ben-Shlomo, Y., Davey Smith, G., Shipley, M., Marmot, M.G.: Magnitude and causes of motality differences between married and unmarried men. Journal of Epidemiology and Community Health 47, 200–205 (1993)
11. Tucker, J.S., Friedman, H.S., Wingard, D.L., Schwartz, J.: Marital history at midlife as a predictor of longevity: alternative explanations to the protective effect of marriage. Health Psychol. 15, 94–101 (1996)
12. Lund, R., Holstein, B.E., Osler, M.: Marital history from age 15 to 40 years and subsequent 10-year mortality: a longitudinal study of Danish males born in 1953. International Journal of Epidemiology 33, 389–397 (2004)
13. Kposowa, A.J.: Marital status and suicide in the National Longitudinal Mortality Study. J. Epidemiol. Community Health 54, 254–261 (2000)
14. Smith, J.C., Mercy, J.A., Conn, J.M.: Marital status and the risk of suicide. Am. J. Public Health 78, 78–80 (1988)
15. Cutright, P., Stack, S., Fernquist, R.: Marital status integration, suicide disapproval, and societal integration as explanations of marital status differences in female age-specific suicide rates. Suicide Life Threat Behav. 37(6), 715–724 (2007)
16. Manzoli, L., Villari, P., Pirone, G.M., Boccia, A.: Marital status and mortality in the elderly: A systematic review and meta-analysis. Social Science & Medicine 64, 77–94 (2007)

17. Blanchard, L.T., Gurka, M.J., Blackman, J.A.: Emotional, developmental, and behavioral health of american children and their families: A report from the 2003 National Survey of Children's Health. Pediatrics 117, e1202–e1212 (2006)
18. González, J., Rojas, I., Ortega, J., Pomares, H., Fernández, F.J., Díaz, A.: Multiobjective evolutionary optimization of the size, shape, and position parameters of radial basis function networks for function approximation. IEEE Transactions on Neural Networks 14(6), 1478–1495 (2003)
19. Guillén, A., González, J., Rojas, I., Pomares, H., Herrera, L.J., Valenzuela, O., Rojas, F.: Output Value-Based Initialization For Radial Basis Function Neural Networks. Neural Processing Letters (June 2007), doi:10.1007/s11063-007-9039-8
20. Guillén, A., Pomares, H., González, J., Rojas, I., Valenzuela, O., Prieto, B.: Parallel multiobjective memetic rbfnns design and feature selection for function approximation problems. Neurocomputing 72(16-18), 3541–3555 (2009)
21. Park, J., Sandberg, J.W.: Universal approximation using radial basis functions network. Neural Computation 3, 246–257 (1991)
22. González, J., Rojas, I., Ortega, J., Pomares, H., Fernández, F.J., Díaz, A.: Multiobjective evolutionary optimization of the size, shape, and position parameters of radial basis function networks for function approximation. IEEE Transactions on Neural Networks 14(6), 1478–1495 (2003)
23. Guillén, A., González, J., Rojas, I., Pomares, H., Herrera, L.J., Valen-zuela, O., Prieto, A.: Improving Clustering Technique for Functional Approximation Problem Using Fuzzy Logic: ICFA algorithm. Neurocomputing (June 2007), doi:10.1016/j.neucom.2006.06.017
24. Guilléen, A., González, J., Rojas, I., Pomares, H., Herrera, L.J., Valenzuela, O., Rojas, F.: Studying Possibility in a Clustering for Function Approximation Algorithm. Neural Computing & Applications (2007)