Memetic Pareto Differential Evolutionary Neural Network for Donor-Recipient Matching in Liver Transplantation

M. Cruz-Ramírez[1,*], C. Hervás-Martínez[1], P.A. Gutiérrez[1], J. Briceño[2], and M. de la Mata[2]

[1] Dept. of Computer Science and Numerical Analysis, University of Córdoba, Spain
[2] Liver Transplantation Unit, Hospital Reina Sofía, Córdoba, Spain, CIBERHED

Abstract. Donor-Recipient matching constitutes a complex scenario not easily modelable. The risk of subjectivity and the likelihood of falling into error must not be underestimated. Computational tools for decision-making process in liver transplantation can be useful, despite its inherent complexity. Therefore, a Multi-Objective Evolutionary Algorithm and various techniques of selection of individuals are used in this paper to obtain Artificial Neural Network models to assist in making decisions. Thus, the experts will have a mathematical value that enables them to make a right decision without deleting the principles of justice, efficiency and equity.

1 Introduction

Liver transplantation is an accepted treatment for patients with end-stage chronic liver disease. Numerous donor and recipient risk factors interact to influence the probability of survival at 3 months after liver transplantation. It is critical to balance waitlist mortality against posttransplant mortality. Our objective was to devise a scoring system that predicts recipient survival at 3 months following liver transplantation to complement the model for end-stage liver disease score (MELD) to predicted waitlist mortality.

Most current organ allocation systems are based on the principle that the sickest patients should be treated first. Models have been developed to estimate the risk of death, considering the underlying disease and urgency of the receiving patient assuming that all donor livers carry the same risk of failure. This, however, is not the case: it has been shown in recent years that the risk of graft failure, and even patient death, after transplantation differs among recipients. While some patients may "tolerate" and overcome the initial poor function of a compromised donor organ, others may not have the same reserve. Increasing

* Corresponding author at: Tel.: +34 957 218 349; Fax: +34 957 218 630; E-mail: mcruz@uco.es M. Cruz-Ramírez's research has been subsidized by the FPU Predoctoral Program (spanish Ministry of Education and Science), grant reference AP2009-0487.

J. Cabestany, I. Rojas, and G. Joya (Eds.): IWANN 2011, Part II, LNCS 6692, pp. 129–136, 2011.

awareness of the diversity in donor organ quality has stimulated the debate on matching between specific recipient and donor factors to avoid futility, but also to avoid personal and institutional differences in organ acceptance. The insufficient supply of deceased donor livers for transplantation has motivated the expansion of acceptance criteria; such organs are captured by the terms marginal and expanded criteria livers. This context of aggressive liver utilization motivated the derivation of the donor risk index, a quantitative, objective, and continuous metric of liver quality based on factors known or knowable at the time of an organ offer.

Thus, predicting the survival of liver transplant patients has the potential to play a critical role in understanding and improving the matching procedure between the recipient and graft. Although voluminous data related to the transplantation procedures is being collected and stored, only a small subset of the predictive factors has been used in modeling liver transplantation outcomes. The previous studies have mainly focused on applying statistical techniques to a small set of factors selected by the domain-experts in order to reveal the simple linear relationships between the factors and survival. Machine learning and soft computing methods offer significant advantages over conventional statistical techniques in dealing with the latter's limitations such as normality assumption of observations, independence of observations from each other, and linearity of the relationship between the observations and the output measure(s). Among these techniques, we will use Artificial Neural Network models (ANN). The use of ANNs in biomedicine as an alternative to other classification methods is baised on different approaches: a Fisher transformation [2,9], due to their flexibility and high degree of accuracy to fit to biomedical data, sigmoid functions and other types of basis functions (Multilayer Perceptron type networks) [11]. In the field of transplantation, ANNs have been designed to diagnose cytomegalovirus (CMV) disease [17] and acute rejection using data obtained from posttransplantation renal biopsies [10] after kidney transplantation. In addition, the use of ANNs was investigated in the prediction of graft failure [15] on the prediction of liver transplantation outcome [6].

ANNs can be trained with Evolutionary Computation algorithms (EC). This methodology widely used in the last few years to evolve neural-network architectures and weights. This is known as Evolutionary Artificial Neural Networks (EANNs), and it has been used in many applications [13,16]. EANNs provide a more successful platform for optimizing network performance and architecture simultaneously.

In this work, we discuss learning and generalization improvement of classifiers designed using a Multi-Objective Evolutionary learning Algorithm (MOEA) [4] for the determination of survival at 3 months after liver transplantation. The data come from eleven hospitals and we investigate the generation of neural network classifiers that achieve high classification level for each class. The methodology is based on two measures: the correct classification rate or Accuracy (C) and the Minimum Sensitivity (MS) as the minimum of the sensitivities of all classes. The aim of this study is to determine which models obtained with the

MOEA, presented the best results. In order to do this, different methods for selection of individuals and ensembles techniques [14] are used, once the execution of the MOEA is finished.

The paper is organized as follows: Section 2 describes the dataset used; Section 3 shows a description of the methodology used; Section 4 explains the experimental design; Section 5 shows the results obtained, while the conclusions and the future work are outlined in Section 6.

2 Dataset Description

A multi-centric retrospective analysis from 11 Spanish units of liver transplantation was conducted, including all the consecutive liver transplants performed between January 1, 2007, and December 31, 2008. The dataset included all transplant recipients aged 18 years or older. Recipient and donor characteristics were reported at the time of transplant. Patients undergoing partial, split or living donor liver transplantation and patients undergoing combined or multi-visceral transplants were excluded from the study. All patients were followed from the date of transplant until either death, graft loss or the first year after liver transplant. Units of liver transplantation were homogeneously distributed throughout Spain.

16 recipient characteristics, 20 donor characteristics and 3 operative factors were reported for each donor-recipient pair. The end-point variable for artificial neural network modeling was 3-month graft mortality. A total of 1031 liver transplants were initially included. The follow-up period was fulfilled in 1003 liver transplants. 28 cases were excluded because the absence of graft survival data. All losses were well distributed among the participating institutions. Donor or recipients with missing entries were not eliminated but missing values were filled using data imputation techniques. Imputation techniques employed are those commonly used. When the number of non-responses in a variable is less than 1%, we have substituted the value by the average if the variable is continuous and by the mode if the variable is discrete (categorical). In another case, we used polynomial regression models to estimate these values.

3 Methods

3.1 Accuracy and Minimum Sensitivity in Classification Problems

To evaluate a classifier, the machine learning community has traditionally used Correct Classification Rate or Accuracy (C) to measure its default performance. Actually, it suffices to realize that C cannot capture all the different behavioral aspects found in two different classifiers in multi-class problems. For these problems, two performance measures are considered: traditionally-used C, as the number of patterns correctly classified and the Minimum of the Sensitivities of all classes (MS), that is, the lowest percentage of examples correctly predicted as belonging to each class, S_i, with respect to the total number of examples in the corresponding class, $MS = \min\{S_i\}$ (for a more detailed description of these measures, see [8]). This is, we assume the premise that a good classifier should combine a high classification rate level in the testing set with an acceptable level for each class.

In [8], C and MS are presented as objectives that could be positively correlated, but, while this may be true for small values of MS and C, it is not so for values close to 1 on both MS and C, where the objectives are competitive and conflicting. This fact justifies the use of a MOEA for training ANNs optimizing both objectives.

3.2 Pareto Differential Evolution Algorithm

This paper uses the MOEA described in [5] for training ANN with sigmoid basis functions. The next section briefly explains the this algorithm. For more details about the Base Classifier Framework or Fitness Functions, see [5].

In this paper, we use one of the most prominent Multi-Objective Evolutionary Algorithms in the bibliography. This algorithm is the MPDENN (Memetic Pareto Differential Evolution Neural Network) algorithm developed by R. Storn and K. Price in [18], modified by H. Abbass to train neural networks [1] and adapted for C and MS [7]. The fundamental bases of this algorithm are Differential Evolution (DE) and the concept of Pareto dominance.

The main feature of the MPDENN algorithm is the inclusion of a crossover operator together with the mutation operator. The crossover operator is based on a random choice of three parents, where one of them (main parent) is modified using the weighted difference of the two other parents (secondary parents). The child generated by the crossover and mutation operator is included in the population if it dominates the main parent, if it has no relationship with him or if it is the best child of the rejected children. At the beginning of each generation, individuals dominated are eliminated from the population. A generation of the evolutionary process ends when the population has been completed. In three generations of the evolution (the first initially, the second in the middle and the third at the end), a local search algorithm is applied to the most representative individuals of the population. The local search algorithm used by MPDENN algorithm is *iRprop*$^+$ [12] (more details in [5]).

3.3 Automatic Selection Method Used in the Experimentation

Once the execution of MPDENN algorithm ends, various automatic selection methods of individuals are used for each run:

- **MPDENN-E:** It consists of choosing the Pareto upper extreme value in training, that is, the best individual in Entropy (E), because one of the fitness function of the MOEA is E. This method is described in [8].
- **MPDENN-MS:** This technique is similar to the previous one, but selecting the best individual in MS, i.e., the individual at the Pareto lower extreme. This method is described in [8].

- **MPDENN-CC:** This method selects all individuals from the first and second Pareto front obtained with the MPDENN algorithm. This group of individuals is divided into two subgroups by a 2-means algorithm (because there are two objective functions, C and MS). The individual that is closest to the centroid of the upper cluster (cluster that takes the C measure into account) is selected.
- **MPDENN-CMS:** This automatic method works in a similar way to the MPDENN-CC automatic method, but in this case, the individual that is closest to the centroid is selected taking the MS measure into account (lower cluster).

  Individuals selected by MPDENN-CC and MPDENN-CMS are considered the most representative individuals in the population (the fact that these individuals do not have the greatest value in any objective does not mean that they do not generalize well).

  We decided to include the second Pareto front in the clustering process, in order to expand the number of individuals and to increase diversity. In addition, individuals belonging to this front may have a high percentage of classification in generalization because it is a way to avoid over-training. In the extreme case that is only one individual in each of the fronts (there would be only two individuals), each of these individuals will be assigned to a cluster.
- **MPDENN-MV [19]:** *Majority Voting* (MV) is an ensemble technique that uses all individuals in the first Pareto front. With this technique, a pattern belong to the class that has the higher number of votes, according to the independent classification of each of the elements that make up the ensemble. To estimate the a posteriori probability of a pattern to belong to a class, the average of the output probabilities of the models who voted for this class are used. This is performed for each pattern in the training of generalization dataset so that a probability matrix is formed to obtain the $RMSE$ measure (Root Mean Square Error).
- **MPDENN-SA [19]:** The *Simple Averaging* (SA) ensemble technique uses the first Pareto front to calculate for each pattern the arithmetic mean of the probability for each Q class for each of the models in the ensemble. The assignment will take the class that has the highest average probability. For the case of the $RMSE$ measure, the arithmetic mean of the probabilities is obtained for each output of each model in the ensemble for a particular pattern. Then we use the probabilities of the output with the maximum mean probability for each model of the ensemble for that particular pattern. This is done for each pattern in the training and generalization dataset, and a probability matrix is formed to obtain the $RMSE$ measure.
- **MPDENN-WT [19]:** With the *Winner Take All* (WT) ensemble method, for each pattern the probabilities of the model with the highest probability in one of the outputs are used as the output of the ensemble. This ensemble method uses the individuals in the first Pareto front.

Table 1. Features of the dataset

#Patterns	#Training patterns	#Test patterns	#Input variables	#Classes	#Patterns per class
1003	751	252	64	2	(890,113)

4 Experimental Study

The Experimental design was conducted using a stratified holdout procedure with 30 runs, where approximately 75% of the patterns were randomly selected for the training set and the remaining 25% for the test set. During the creation of these two sets, the proportion of 75-25% for the training-testing patterns for each of the participating hospotals was also kept. Table 1 shows the features of the dataset.

In all the experiments, the population size for MPDENN is established as $M = 25$. The crossover probability is 0.8 and the mutation probability is 0.1. For *iRprop*$^+$ as local search algorithm, the adopted parameters are $\eta^+ = 1.2$, $\eta^- = 0.5$, $\Delta_0 = 0.0125$ (the initial value of the Δ_{ij}), $\Delta_{\min} = 0$, $\Delta_{\max} = 50$ and $Epochs = 10$, see [12] for the *iRprop*$^+$ parameter description.

To start processing data, each one of the input variables was scaled in the ranks $[-1.0, 1.0]$ to avoid the saturation of the signal. Addition, categorical variables have been transformed into many binary variables as possible category.

5 Results

The C and the $RMSE$ represent two most often used metrics in classification [3]. In our paper, we use these two metrics together with the MS. Table 2 presents the values of mean and Standard Deviation (SD) for C, MS and $RMSE$ in generalization in 30 runs of all the experiments performed. The analysis of the results leads us to conclude that the MPDENN-E obtained the best performance in the dataset considering C_G and the second best value in $RMSE_G$. The best result in $RMSE_G$ is obtained by the MPDENN-CC. For MS_G, the MPDENN-MS obtained the best results in the analyzed dataset. From this analysis, we can consider that the best method for C_G and $RMSE_G$ is the MPDENN-E while it is MPDENN-MS for MS_G.

The best models obtained by MPDENN-E and MPDENN-MS methods are shown in Table 2. The best MPDENN-E model has a high value on $C = 89.29$ and $RMSE = 0.3212$, while the best MPDENN-MS model produces a very acceptable value on $MS = 62.07$. The confusion matrix for the best MPDENN-E model in generalization is: True Positive (TP)=221, False Negative (FN)=2, False Positive (FP)=25, True Negative (TN)=4; and for the best MPDENN-MS model is: TP=143, FN=80, FP=11, TN=18.

These results suggest that a combination of the two models extremes of the Pareto front would provide a useful tool for the problem of donor-recipient assignment. This combination could be a rule-based system or a weighted aggregation

Table 2. Statistical results for different methods in generalization

Method	$C_G(\%)$ Mean ± SD	$MS_G(\%)$ Mean ± SD	$RMSE_G$ Mean ± SD
MPDENN-E	**88.34 ± 0.68**	1.15 ± 3.98	*0.3282 ± 0.0071*
MPDENN-MS	60.66 ± 3.04	**50.39 ± 6.56**	0.4492 ± 0.0763
MPDENN-CC	88.24 ± 0.66	1.03 ± 4.54	**0.3261 ± 0.0068**
MPDENN-CMS	60.71 ± 3.14	*50.23 ± 8.48*	0.4160 ± 0.0561
MPDENN-MV	68.33 ± 8.49	27.35 ± 11.11	0.3583 ± 0.0184
MPDENN-SA	84.26 ± 3.87	8.97 ± 7.72	0.3454 ± 0.0122
MPDENN-WT	*88.25 ± 0.67*	0.69 ± 2.29	0.3581 ± 0.0733
Method	$C_G(\%)$	$MS_G(\%)$	$RMSE_G$
Best MPDENN-E model	89.29	13.79	0.3212
Best MPDENN-MS model	63.89	62.07	0.3863

The best result is in **bold** face and the second best result in *italics*.

of the outputs of both models, although in our opinion, the rules-based system would provide a more understandable and comprehensible tool for experts. The system would receive as input a set of potential recipients and form a donor-recipient pair between each of them and donor/organ data. These pairs would be the input for these neural network models. With the results provided by these models and using a simple set of rules, the system would determine which of the recipients receiving the organ.

6 Conclusions

With the study presented in this paper, we obtain some artificial neural networks models that can help medical experts in the donor-recipient allocation. These models are obtained by a multi-objective evolutionary algorithm where Accuracy is the measure considered to evaluate model performance along with the Minimum Sensitivity measure. Minimum Sensitivity is used to avoid the design of models with high global performance but bad performance when considering the classification rate for each class (survival or not-survival).

With the two best models (obtained by MPDENN-E and MPDENN-MS methods), a rule-based system can be used to perform the matching between donor and recipient. This rule-based system must be generated by the medical experts and machine learning experts, for maintaining the principles of justice, efficiency and equity. The current allocations systems are based on thee risk of death on the waiting list and do not recognize distinctions in “donor organ quality”. With the rule-based system, the “donor organ quality” would be taken into account to improve the allocation and ensure the survival of recipients.

References

1. Abbass, H.A., Sarker, R., Newton, C.: PDE: a Pareto-frontier differential evolution approach for multi-objective optimization problems. In: Proceedings of the 2001 Congress on Evolutionary Computation, Seoul, South Korea, vol. 2 (2001)

2. Bishop, C.M.: Neural networks for pattern recognition. Oxford University Press, Oxford (1996)
3. Caruana, R., Niculescu-Mizil, A.: Data mining in metric space: An empirical analysis of supervised learning performance criteria, pp. 69–78 (2004)
4. Coello Coello, C., Lamont, G., Veldhuizen, D.: Evolutionary Algorithms for Solving Multi-Objective Problems, 2nd edn. Springer, Heidelberg (September 2007)
5. Cruz-Ramírez, M., Sánchez-Monedero, J., Fernández-Navarro, F., Fernández, J., Hervás-Martínez, C.: Memetic pareto differential evolutionary artificial neural networks to determine growth multi-classes in predictive microbiology. Evolutionary Intelligence 3(3-4), 187–199 (2010)
6. Dvorchik, I., Subotin, M., Marsh, W., McMichael, J., Fung, J.: Performance of multi-layer feedforward neural networks to predict liver transplantation outcome. Methods Inf. Med. 35, 12–18 (1996)
7. Fernández, J.C., Hervás, C., Martínez, F.J., Gutiérrez, P.A., Cruz, M.: Memetic pareto differential evolution for designing artificial neural networks in multiclassification problems using cross-entropy versus sensitivity. In: Corchado, E., Wu, X., Oja, E., Herrero, Á., Baruque, B. (eds.) HAIS 2009. LNCS, vol. 5572, pp. 433–441. Springer, Heidelberg (2009)
8. Fernández, J.C., Martínez-Estudillo, F.J., Hervás-Martínez, C., Gutiérrez, P.A.: Sensitivity versus accuracy in multiclass problems using memetic Pareto evolutionary neural networks. IEEE Trans. on Neural Networks 21(5), 750–770 (2010)
9. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Annals of Eugenics 7(7), 179–188 (1936)
10. Furness, P.N., Levesley, J., Luo, Z., Taub, N., Kazi, J., Bates, W., Nicholson, M.: A neural network approach to the biopsy diagnosis of early acute renal transplant rejection. Histopathology 35(5), 461–467 (1999)
11. Haykin, S.: Neural Networks: A comprehensive Foundation, 2nd edn. Prentice Hall, Upper Saddle River (1998)
12. Igel, C., Hüsken, M.: Empirical evaluation of the improved rprop learning algorithms. Neurocomputing 50(6), 105–123 (2003)
13. Kondo, T.: Evolutionary design and behavior analysis of neuromodulatory neural networks for mobile robots control. Appl. Soft Comput. 7, 189–202 (2007)
14. Löfström, T., Johansson, U., Boström, H.: Ensemble member selection using multiobjective optimization. In: IEEE Symposium on Computational Intelligence and Data Mining, pp. 245–251 (2009)
15. Matis, S., Doyle, H., Marino, I., Mural, R., Uberbacher, E.: Use of neural networks for prediction of graft failure following liver transplantation. In: IEEE Symposium on Computer-Based Medical Systems, pp. 133–140 (1995)
16. Saxena, A., Saad, A.: Evolving an artificial neural network classifier for condition monitoring of rotating mechanical systems. Appl. Soft Comput. 7, 441–454 (2007)
17. Sheppard, D., McPhee, D., Darke, C., Shrethra, B., Moore, R., Jurewitz, A., Gray, A.: Predicting cytomegalovirus disease after renal transplantation: an artificial neural network approach. Int. J. Med. Inf. 54(1), 55–76 (1999)
18. Storn, R., Price, K.: Differential evolution. A fast and efficient heuristic for global optimization over continuous spaces. J. of Global Optimization 11, 341–359 (1997)
19. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 3rd edn. Elsevier, Academic Press (2006)