ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 14, No. 4, 2019, *pp*.266-271

Optimization of Atmospheric Plasma Surface Modification Process Using Decision Trees

Radhia Abd Jelil [1]

[1] *Textile Material and Processes Research Unit, University of Monastir, Tunisia*

(Received October 11 2019, accepted November 28 2019)

Abstract. Decisions trees are one of the most commonly used data mining techniques to practically solve classification and prediction problems. They have tree shaped structures in which construction of trees is simple and unlike the logistic regression models, decision tree results can be easily understood by the users. In this study, a decision tree induction algorithm known as CART (Classification and Regression Trees) has been employed in order to better understand the influence of plasma parameters adjustment on polypropylene (PP) film's hydrophilic surface properties. The cross-validation method was used for pruning the decision tree. The root mean square errors (RMSE) and correlation coefficients (R) for training and test subsets were used in order to get the best fitting model. The obtained decision tree regression model showed excellent learning performance and achieved good predictive accuracy.

Keywords: Atmospheric plasma process, polypropylene, optimization, decision trees

1. Introduction

Surface treatment of textile and polymeric surfaces is usually necessary to alter their surface characteristics and to improve their adhesion properties. Many surface modification techniques like wet-chemical treatment, UV irradiation, and plasma treatment have been applied to polymer films to enhance their hydrophilic properties. Among these technologies, atmospheric-pressure plasma treatment has attracted much attention due to its dry process, low operation cost, and high productivity. This technology has been recognized as an environmentally friendly alternative to conventional wet-chemical processes, since it does not require the use of water and chemicals and since there no waste production [1].

A plasma is a partially ionized gas in neutral state containing highly excited atomic, molecular, ionic and radical species, as well as photons and electrons. During plasma treatment, these energetic species interact strongly with the substrate surface, usually via free-radical chemistry. The interaction depth is confined only to a few ten of nanometers without impairment of the material bulk properties. Essentially, four major effects on surfaces can be obtained depending on treatment conditions, namely cleaning, ablation, cross-linking, and surface activation. Many studies have been undertaken to investigate the effect of plasma treatment on polypropylene surfaces [2-5]. Though plasma treatment may result in many desirable surface modifications, there has not been much effort to optimize the main set of parameters governing plasma process, and then reliably reproduce the process outcome. Therefore, the exploration of data mining techniques like decision trees seems to be promising. Decision trees are a non-parametric supervised learning method that can be used for classification and regression. They are simple in construction, easy to understand, and robust even in the presence of missing data [6-8]. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. The main components of a decision tree model are nodes, branches, and leaf nodes (terminal nodes). Each node represents a mathematical or logical test upon specific attributes in the data set. Parent nodes can have two or more child nodes, depending on the induction algorithm chosen. The parent and child nodes are connected via branches that represent the outcomes of the test performed at the parent node. A leaf node has no children and represents a class label (decision taken after computing all attributes). The decision tree pathways can also be represented as "if-then" rules.

The objective of this study was to build a prediction model, using decision tree, for predicting polypropylene surface modification by atmospheric air plasma. A set of experimental data was collected for training this model. The inputs variables of the model were plasma treatment power and treatment speed. The target variable was polypropylene surface tension.

The remainder of this paper is organized as follows: Section 2 describes the material and methods used. Section 3 presents and discusses the results, and finally Section 4 concludes the paper.

2. Material and methods

2.1. Plasma treatment

The substrate used in this study was a polypropylene film of a 30 µm-thick. This film was put on the atmospheric plasma machine called "Coating Start" manufactured by Ahlbrant System (Germany) to carry out atmospheric air plasma treatment [9]. The following machine parameters were kept constant: frequency of 26 KHz, electrode length of 0.5 mm and inter-electrode distance of 1.5 mm. The process factors that were varied include the electrical power and treatment speed. Their experimental ranges are shown in Table 1. The PP surface tension data was collected by means of contact angle measurement. A total of 16 samples were performed to develop our predictive decision tree model. Experimentally, we have found that the surface tension of PP film increases to some extent with increasing electrical power and with decreasing the treatment speed.

Table1. Experimental factors and ranges.

Parameter	Range	Unit
Electrical power	300-1000	Watts
Treatment speed	2-10	m/min

2.2. Decision tree approach

In the present study, decisions trees were constructed using the CART algorithm. CART stands for classification and regression trees, a non-parametric statistical algorithm developed by Leo Breiman et al. [10]. CART is a binary recursive partitioning technique. CART methodology comprises three main stages: growing or splitting decision trees, pruning, and selection of the optimal tree.

2.2.1. Splitting. The process of tree building begins by splitting the root node into two child nodes. CART computes the best split by considering all probable splits for each independent or explanatory variable. The best split is obtained when the impurity function, which exists between the parent node and two child nodes, is minimized. Using best split, which reduces impurity as a splitting criterion, an over large or complex tree is grown following recursive partitioning of the nodes. Though the tree interprets data perfectly, when it overfits the data, the predictive ability becomes low. Therefore, there is a need to build a tree with better accuracy and predictive ability.

2.2.2. Pruning. Pruning develops an optimal tree, by shedding off the branches of the large tree. The pruning procedure develops a sequence of smaller trees and computes cost complexity for each tree. The cost complexity is measured by the number of leaves in the tree, and error rate of the tree. Based on the cost-complexity parameter, the pruning procedure determines the optimal tree with high accuracy.

2.3.3. Selection of the optimal tree. The optimal tree is one that has the smallest prediction error for new samples. In our case, prediction error is measured using cross-validation. This method consists in dividing the sample in 10 groups or 'folds', and testing the model developed from 9 folds on the 10th fold, repeated for all ten combinations, and averaging the rates or erroneous predictions. Among the different trees, the simplest tree that has the lowest cross validation error rate is selected as the optimal tree. But the tree at this stage is considered to be unstable, as the results predicted from the decision tree can change rapidly with slight change in the training dataset. Therefore, in order to avoid instability, a "one-standard error" rule is used with cross-validation. According to this rule, the tree with the smallest size and cross-validation error within one standard deviation error of the minimal cross-validation error is selected as the optimal tree. The selection of the optimal tree concludes the final stage of CART and the prediction is made by observing the tree from the root node to the terminal nodes and values in the terminal nodes. The performance of the predictive model is assessed by means of Pearson's correlation coefficient (R) and root mean square error (RMSE). The RMSE gives the standard deviation of the model prediction error, so the smaller its value, the better the model performance is.

3. Results and discussions

The decision tree induction algorithm (CART) was implemented with the help of the MATLAB software. In a first stage, a complex tree with maximum size was grown by recursively partitioning the data. Figure 1 illustrates the fully induced decision tree.

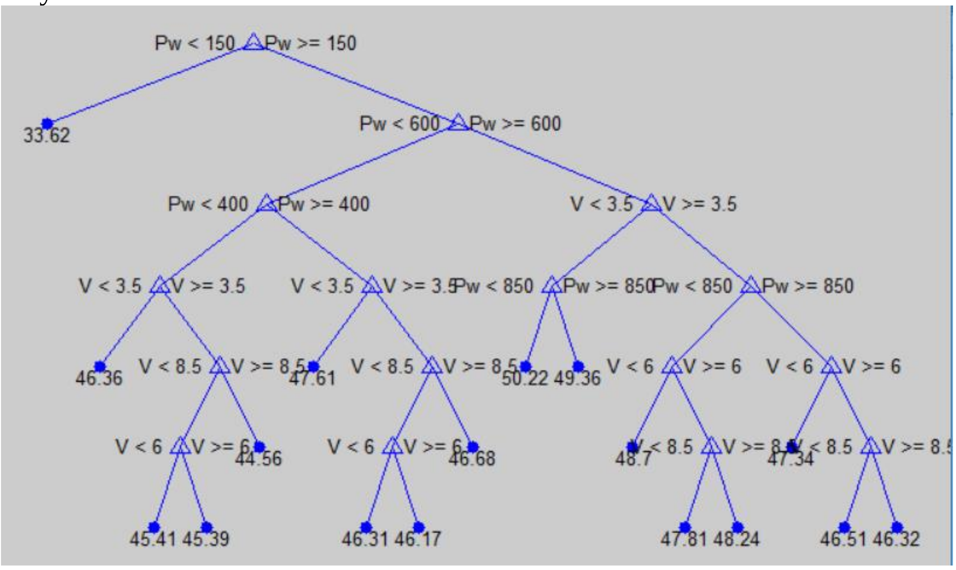

Fig.1. Fully induced decision tree.

It can be seen that the obtained decision tree has many terminal nodes. Though this tree interprets data perfectly, when it over-fits the data, the data predictive ability becomes low. Therefore, 10-fold cross-validation method is used in order to determine the optimal level of tree complexity, and thus improving model accuracy. In a second stage, the fully induced decision tree is pruned back slightly further than the cross-validation minimum error. Figure 2, shows the training and cross-validation errors of CART model as a function of the tree size (Number of terminal nodes). According to this figure, the tree of size 5 is selected by cross-validation as the tree that best fits the data. Hence, the original decision tree is pruned to obtain an optimal tree which has 5 terminal nodes, as shown in figure 3.

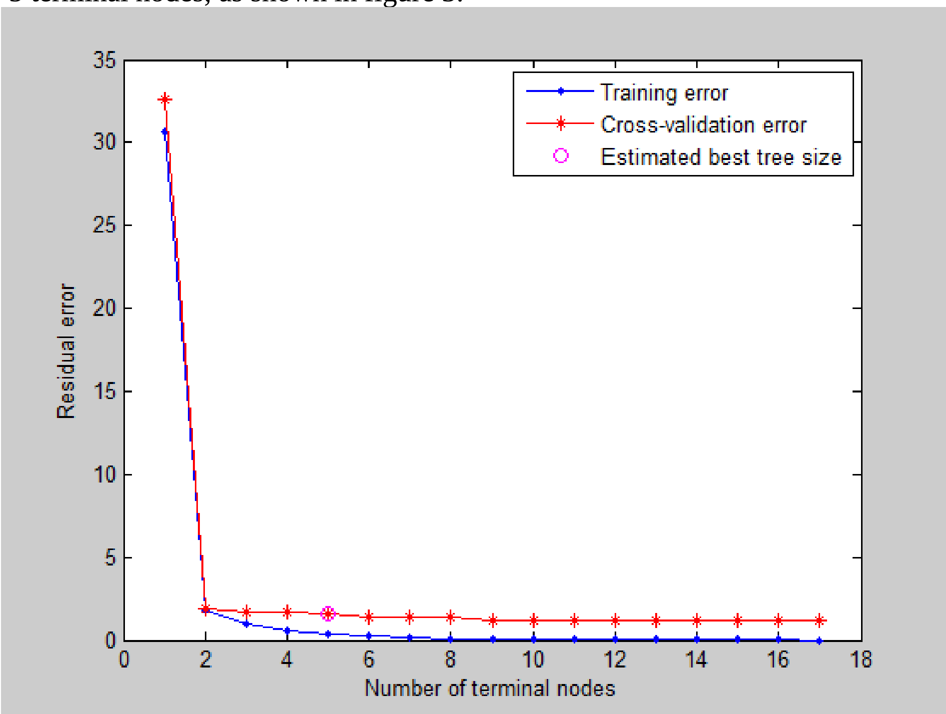

Fig.2. Evolution of decision tree training and cross-validation errors versus tree size.

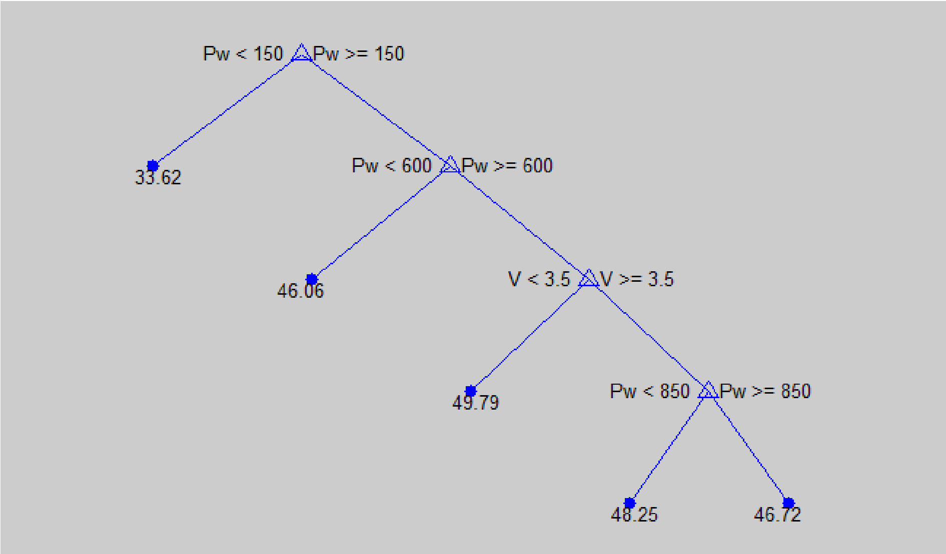

Fig.3. Optimal decision tree.

The last stage of decision tree construction involves extracting decision rules. Each path from the root of a decision tree to one of its leaves can be transformed into a rule simply by conjoining the tests along the path to form the antecedent part, and taking the leaf's class prediction as the class value. In our case, the pruned tree has 5 leaves. So, we can easily deduce 5 rules (Table 2).

Table 2. Rules extracted from the tree of Figure 3.

N°	Premise	Conclusion
1	Electrical power < 150	Surface Tension = 33.62
2	Electrical power≥ 150 and < 600	Surface tension = 46.06
3	Electrical power ≥ 600, and Treatment speed< 3.5	Surface tension = 49.79
4	Electrical power≥ 600 and < 850, and Treatment speed≥ 3.5	Surface tension = 48.25
5	Electrical power ≥ 850, and Treatment speed≥ 3.5	Surface tension = 46.72

These extracted rules are useful to better understand the influence of plasma parameters adjustment on polypropylene (PP) film's hydrophilic surface properties. Therefore, decision trees can provide deep insights into the plasma surface modification process. The overall model performance is demonstrated by the RMSE on the training and test subsets which are respectively equal to 0.46 and 0.96. Although promising, it is often useful to perform a correlation analysis in order to get a true unbiased indication of the decision tree regression model's performance. Figures 4 and 5 exhibit the scattering diagrams of the obtained model.

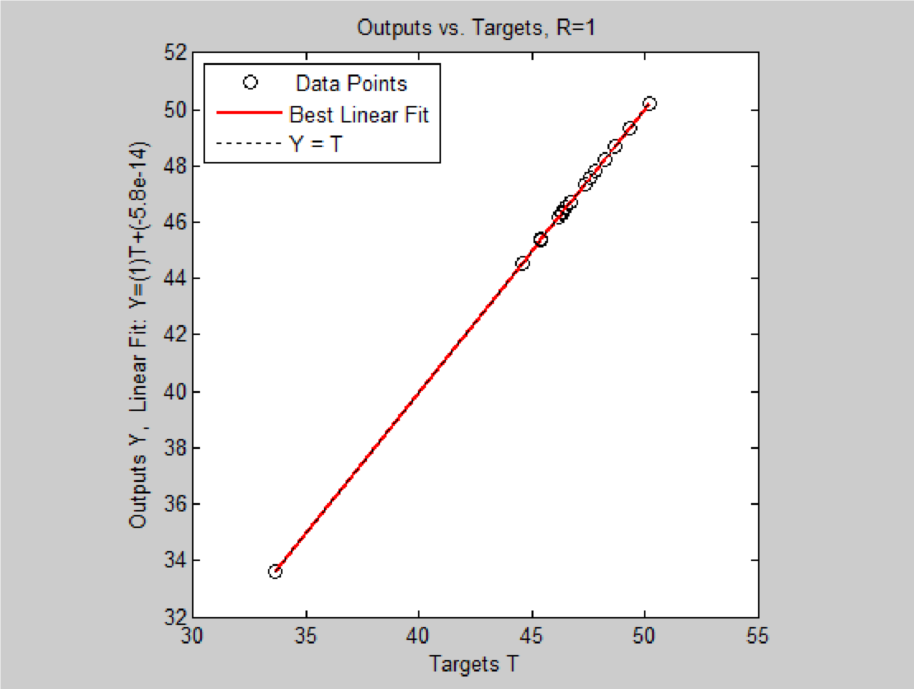

Fig.4. Correlation between output and target values over the training subset.

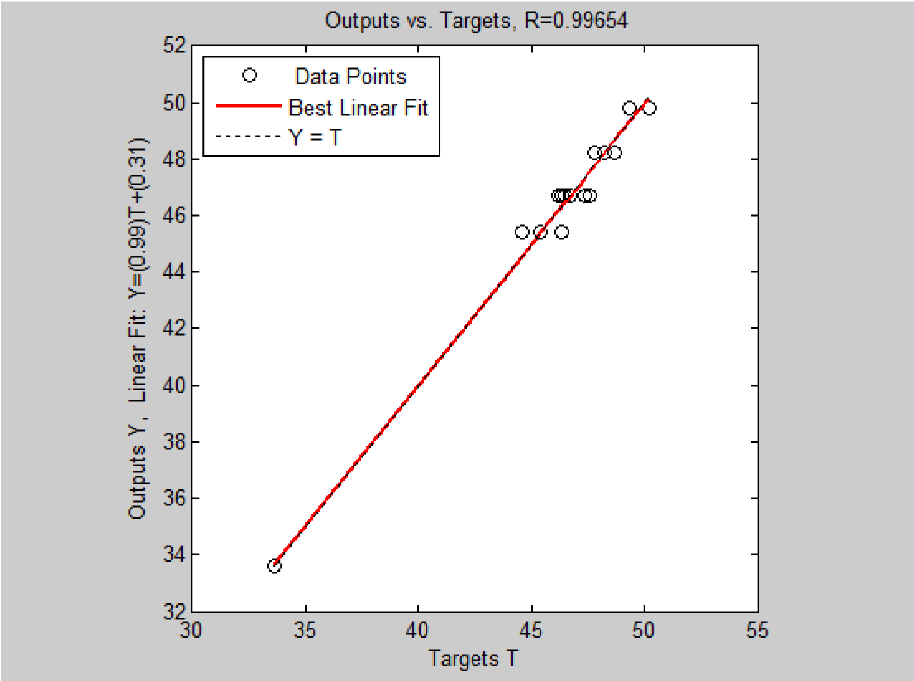

Fig.4. Correlation between output and target values over the test subset.

It can be clearly seen that the linear coefficient of correlation is very high between observed experimental data and CART predicted values. The R-value is 1 in training and 0.9965 during test phase. Thus, it can be deduced that the decision tree regression model has successfully modeled the current data and it is able to make prediction based on the currents input provided. Furthermore, CART provides valuable information regarding the influence of the different input parameters, since the structure of the tree represents the regression process until the result is reached. Electrical power showed up as an important input parameter for an accurate prediction of the output variable. Indeed, electrical power relates to plasma energy which significantly influences the change in surface tension.

4. Conclusion

In this study, a decision tree regression model about modifying polypropylene surface tension has been constructed using CART algorithm. The obtained results showed that the developed model exhibited high correlation coefficients and acceptable prediction errors, showing that it learning and prediction performances were good enough. Thus a conclusion can be drawn that decision tree regression model was capable of describing the relationship between the plasma processing parameters and surface tension properties. Nevertheless, it should be mentioned that decision tree regression model would give better results if the training data set covered greater operating range of process variables.

References

[1] R. A. Jelil, *A review of low-temperature plasma treatment of textile materials*, J. Mater. Sci. 50 (2015), pp.5913-5943.

[2] M. G. McCord, Y. J. Hwang, P. J. Hauser, Y. Qiu, J. J. Cuomo, O.E. Hankins, M. A. Bourham, and L. K. Canup, *Modifying nylon and polypropylene fabrics with atmospheric pressure plasmas*, Text. Res. J. 72 (2002), pp.491-498.

[3] Y. J. Hwang, M. G. McCord, J. S. An, B. C. Kang, and S. W. Park, *Effects of helium atmospheric pressure plasma treatment on low-stress mechanical properties of polypropylene non-woven fabrics,* Text. Res. J. 75 (2005), pp.771-778.

[4] C. H. Su, T. H. Chen, S. H. Yang, C. H. Liu, S. Lin, J. T. Teng, and H. Chen, *Surface properties of polypropylene treated using atmospheric pressure plasma jet*, In: Hinduja S., Fan K. C. (eds). Proceedings of the 35th International MATADOR Conference (2007), Springer, London.

[5] M-H. Chiang, K-H. Liao, I-M. Lin, C-C. Lu, H-Y. Huang, C-L. Kuo, and J-S. Wu, *Modification of Hydrophilic Property of Polypropylene Films by a Parallel-Plate Nitrogen-Based Dielectric Barrier Discharge Jet*, IEEE T. Plasma Sci. 38 (2010), pp.1489 – 1498.

[6] Y-y. Song, Y. Lu, *Decision tree methods: applications for classification and prediction*, Shanghai Arch. Psychiatry, 27 (2015), pp.130-135.

[7] B. Bakir, I. Batmaz, F. A. Gunturkun, I. A. Ipekçi, G. Koksal, and N. E. Ozdemirel, *Defect cause modeling with decision tree and regression analysis*, WASET, 24 (2008), pp.821-824.

[8] C. Liu, Z. H. Rather, Z. Chen, and C. L. Bak, *An overview of decision tree applied to power systems*, Int. J. Smart Grid Clean Energy, 2 (2013), pp.413-419.

[9] R. A. Jelil, Z. Xianyi, L. Koehl, and A. Perwuelz, *Modeling plasma fabric surface treatment using fuzzy logic and artificial neural network*, JIC, 8 (2013), pp.141-152.

[10] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*, Chapman & Hall/CRC, 1984.