

# Knowledge Discovery System For Cost-Effective Composite Polymer Selection-Data Mining Approach

**Doreswamy**

Dept. of Computer Science,  
Mangalore University,  
Mangalagangothri-574 199,  
Karnataka, INDIA.  
Email:doreswamyh@yahoo.com

**S. C. Sharma**

CMRTU, R.V Engineering  
College, Bangalore  
Karnataka, INDIA  
Email: rvrdses@blr.vsnl.net.in

**M. Krishna**

CMRTU, R.V. Engineering  
College, Bangalore  
Karnataka, INDIA  
Email: rvrdses@blr.vsnl.net.in

## Abstract

In this paper Knowledge Discovery System through Data Mining technique is implemented for discovering the knowledge of cost effective polymer composites that meet the design requirements of a component. The Decision Tree Classifier (DTC) with linear function classifies fiber class into short, medium and long fiber classes on fiber critical length  $l_c$ . Appropriate quantifiable characteristics as such, volume fractions, volumes, masses and costs of long fibers and of polymer matrix are probed through the composite rule mixture formula. The knowledge about a reinforcement fiber for a given polymer matrix is discovered through the examined parameters. The knowledge discovered from this system will be useful for design engineering applications where a cluster of handful calculations involved for the least cost composite polymer selection. A set of design specifications for a polymer composite cylindrical pipe with longitudinal filament winding is computed and fed as input parameters for examining the knowledge discovery system.

### Keywords:

Decision Tree Classifier, Linear function, Knowledge Discovery System, Data Mining, Polymer Matrix, Reinforcement fiber, Composite Rule Mixture formula.

## 1. Introduction

It is estimated that there are more than 50,000 useful metallic and non-metallic engineering materials, included with metals, ceramics, polymers, composites, reinforcements, glasses, carbons, fibers and semiconductors, currently available for material processing technology.

This large set of materials and the manufacturing process available to engineers, coupled with complex relationships between the different quantifiable characteristics, often make selection of materials for a given component a difficult task. If the selection process is made haphazardly, there will be a risk of overlapping possible attractive and alternative materials. The risk can be reduced by designing expert systems with artificial intelligent computational algorithms that are effective for decision making on large data set. A variety of computer-aided expert systems have been designed using quantitative procedures for material selection process [1-5], but their limitations are confined to material selection but not for the reinforcements classification and selection.

Polymer processing technology is a branch of material processing technology that replaces the conventional materials with equivalent polymeric materials in many diverged areas such as Computer, Marine, Medical, and Aerospace technologies. The plastic designers and manufacturers are the major components of this technology and most plastic designers select Polymer Matrix(PM) and Reinforcement Fibers(RF) with the aid of materials handbooks such as GE's "Design Guidelines", Mobays's "Design Manuals" and DuPont's "Design Handbooks". This type of resource generally provides an overview of PMs and RFs materials followed by a detailed portrayal on each specific material and its properties. These handbooks tend to be very extensive but often non-intuitive.

Therefore, one must know the particular type of PM and RF material needed and do enormous mathematical computations on specific quantitative properties to meet the design constraints.

Some companies have accelerated this process by locating the materials' information in computerized databases. Engineering Design Database (EDD)[6] and Mobay's PC-Based CAMPUS plastic database [11] typically feature a search mechanism to assist

designers in selecting candidate materials. The selection methodology of these is one of the initial screening processes [9] for the best possible match for specific quantitative properties.

There is an elegant possibility to select polymer composites with an aid of Graphical Techniques [6-12] that give general idea about PMs and RFs. This technique is limited to two quantitative properties of materials. The task of selection process is made more complex with GTs if more than two properties are considered to measure the performance of composite materials. Hence, Knowledge Based Expert Systems are proposed since they are capable for handling and computing large volume of data involved in making effective decisions on multi-attributes.

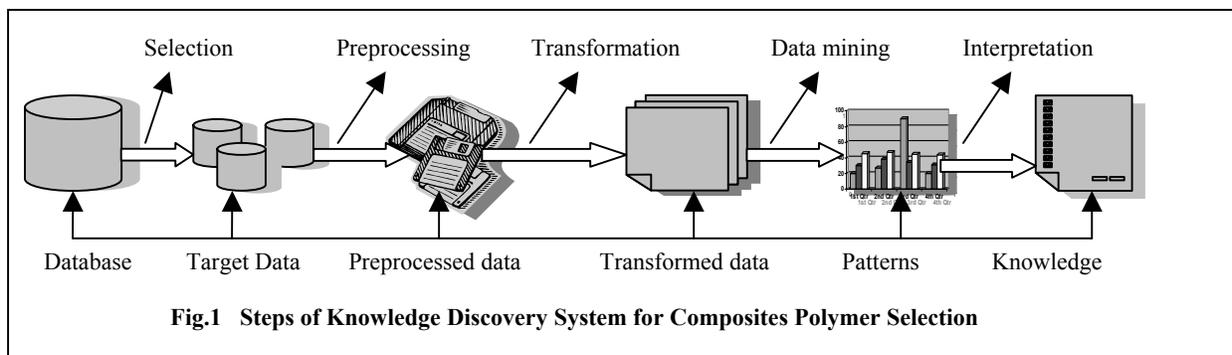
This paper is organized as follows. The 2<sup>nd</sup> section describes the selection; preprocessing and transformation steps of KDS. Decision Tree Classifier for fiber classification is presented in section 3. Exploratory analysis and composite rule mixture computation and with classifier algorithm are

depicted in section 4. The results and conclusion are given section 5.

## 2. The Knowledge Discovery System (KDS)

The proliferation of large masses of data has created many new opportunities for those who are working in Scientific, Engineering and Business fields. The opportunities are offered by the abundance and availability of data and the challenges are posed by the problems of how to organize, retrieve and extract knowledge from the database [13, 14]. The complex problems involving retrieval of knowledge from the large data space are easily solved by the KDSs.

The KDS is interactive and iterative, involving numerous steps with many decisions made by the user. However, when we engage in real discovery tasks, we find that these tasks can be very complex. Brachnman and Anand[14] have identified all of the steps of KDS as realistic easy segments and these shown in the fig.1.



### 2.1. Understanding Domain Knowledge

Developing an understanding of the application domain, the relevant prior knowledge and identifying the goals of the Knowledge Discovery System (KDS) from the customers' view point is the primary step that resolves the complex steps into easier steps in discovering user understandable knowledge. Deriving knowledge of polymer composites that satisfies the design requirements is a critical task as the database consisting of vast materials properties in huge numbers. Consequently selection problem is multi-attribute decision making problem where each attribute corresponds to one of the product profile requirements.

A tubular filament-wound composites shaft is to be structured that has an output diameter of 70mm, inside diameter of 58.8mm and a length 1.22m. The mechanical characteristic of prime importance is bending stiffness in terms of longitudinal modulus of elasticity ( $E_c$ ). Stiffness is to be specified as maximum

allowable deflection; when subjected to three-point bending, a load 890N is to produce elastic deflection not more than 0.33mm at the midpoint position. A circumferential winding pattern is to be used with maximum winding pitch of the fiber expressed in  $\theta$ . The maximum stiffness in the longitudinal direction in this design is  $\theta = 15^\circ$ . The decision constraints for the desired knowledge are as follows.

1. Decide the fiber material, when embedded in the polymer matrix, meets the stipulated criteria.
2. Select the fiber material that will yield the lowest cost filament-wound composite material.

Material attributes critical to selection are categorized as mechanical and physical properties.

Mechanical Property - Modulus of Elasticity ( $E_f$ )GPa

Physical Property - Density ( $\text{g/cm}^3$ ).

General property - Cost/Kg.

The above specifications are the input parameters to the KDSs and assumed that these are given by a design engineer.

## 2.2. Creating a logical data set

The databases of Polymer Matrix and Reinforcement Fibers are organized with a set of tuples containing the attributes having quantitative range values. Each tuple in a database is expressed as follows

$$T \rightarrow \langle A_1, A_2, A_3, A_4 \dots A_m \rangle \quad (1)$$

Where  $A_1 \dots A_m$  are the attributes with boundary values of a material. The databases are segregated into PM and CR target databases, each database contain a tuple as shown in (1).

## 2.3. Data Preprocessing

Basic operations include removing noise of data sets, collecting the necessary information to model or deciding on strategies for handling missing data fields and account for know changes. Polymer data tend to be incomplete and inconsistency as the attribute values depend on the physical, mechanical, chemical impacts on the entities. The attributes normally have range values start from lower to upper boundary limits and specify desirable properties of PMs and RFs. Missing of either a lower/ an upper boundary value is replaced with the current boundary value for building decision strategies on reduced data space. Thus results with preprocessed data set.

## 2.4. Data Reduction and Projection

Finding useful feature to represent the data is depending on the goal of the task. With transformation, the effective number of attributes, having range values (lbv, ubv) under consideration, can be reduced or invariant representation of data can be found. Aggregation (Statistical Mean) operation on the attribute lower boundary (lbv) and upper boundary (ubv) values yields the mean value that replaces two attributes of polymer and composite materials in the databases. The projection operation is a process of pruning database (unwanted attributes). Thus results with pruned databases of PMs and RFs, that are considered for composite selection.

$$A_{avg} = \left( \frac{A_i lbv + A_i ubv}{2} \right) \quad (2)$$

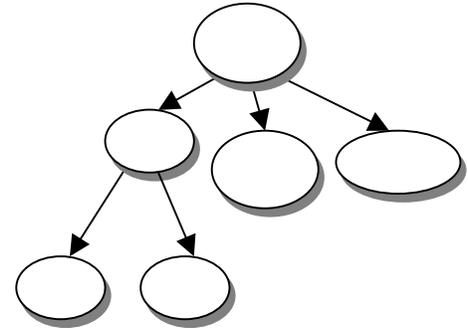
This process obtains the preprocessed data into transformed data that are considered for mining knowledge of composite material.

## 3. Decision Tree Classifier

Matching the goals of KDS is the computational part of Data Mining technique. One such technique implemented here is Decision Tree Classifier

(DTC)[15]. The Decision Tree Classifier is used successfully in many diverse area such signal processing, pattern recognition, remote sensing, medical diagnosis, expert systems and speech recognition. The most important feature of DTC's is its capability to break down a complex decision making process into a collection of simpler decisions. Thus provides a solution which is often easier to interpret.

In the decision tree, each internal node represents a test on an attribute, each branch represents an outcome of the test, and leaf node represents classes or class distribution. It is a divide and conquered approach to the problem of learning from a set of independent test entities. The decision tree is implemented on a test data and sample data sets of reinforcements.



The root node at level 1 defined with discrimination function  $C(D,F)$  on decision class  $D = \{d_1, d_2, d_3, d_4\}$  containing decision rules and the reinforcement fiber data space  $F_{n,m} = \{t_1, t_2, t_3 \dots t_n\}$ , containing  $n$  tuples and each tuple contains  $m$  number of attributes. The discrimination function is linear function  $\varphi(f)$  classifies fiber data space into short(S), medium (M) and long (L) fiber classes. Further it discriminates long fiber class into desired (DF) and undesired fiber (UF) classes. Modulus of elasticity ( $E_f$ ) of long fibers is greater than the expected Modulus Elasticity ( $E_c$ ) of composite. These classifications are based on the fixed critical length  $l_c$  and ( $E_f$ ) of fibers.

The discrimination function  $\varphi(f)$  applied at level 1 for yielding fiber classes L, M and S is defined with fiber length  $L_f$  and Modulus of Elasticity  $E_f$

$$\varphi(f) = \begin{cases} L & \text{if } \frac{(f-E_f)}{15l_c} < 1 \\ M & \text{if } \frac{l_c}{(f-E_f)} < 1 < \frac{15l_c}{(f-E_f)} \\ S & \text{if } (f-E_f) < 1 \end{cases} \quad (3)$$

Further at the level 2, DTC with function  $\varphi(f)$  on fiber class L discriminates into desired and undesired fiber classes labeled with DF and UF respectively. The leaf node DF contains the optimal reinforcement fibers whose length is greater than fifteen times of the critical length ( $l_c$ ) of fiber and  $E_f$  is greater than  $E_c$  of polymer composite. The desired fiber (DF) class is being classified by the following function.

$$\varphi(f) = \begin{cases} DF & \text{if } \frac{(f-L_f)}{E_c} > 1 \\ UF & \text{otherwise} \end{cases} \quad (4)$$

where  $f = W(L_f + E_f)$ ,  $W$  is selection quantitative parameter.

#### 4. Exploratory Analysis on Rule Mixture formula.

It is a core task in KDS for computing appropriate parameters, which have relationship with the output of data mining algorithm, for researching for data patterns that constitute core knowledge. As the design requirement specified in the section 2.1, it is required to compute parameters such longitudinal modulus of elasticity  $E_c$ , Volume  $V_c$  of composite having intra relationship with mid point deflection  $\Delta y$  and a moment of inertia  $I$  of cross sectional area with inner and outer diameters  $d_{in}$  and  $d_o$  respectively.

$$E_c = \frac{4FL^3}{3\pi \Delta y (d_o^4 - d_{in}^4)} \text{ GPa}, \text{ where } \Delta y = \frac{FL^3}{48EI}$$

$$\text{and } I = \frac{\pi}{64} (d_o^4 - d_{in}^4) \quad (5)$$

$$V_c = \frac{\pi L}{4} (d_o^4 - d_{in}^4) \text{ cm}^3 \quad (6)$$

The longitudinal modulus of elasticity  $E_c$  of composite material with long fiber having deflection  $\theta$ , is derived with composite rule mixture formula [16].

$$E_c = \text{Cos}\theta(E_m v_m + E_f v_f) \text{ and } v_m + v_f = 1 \quad (7)$$

The volume of composite  $V_c$ , is proportional to the relative volumes  $V_m$  and  $V_f$  of Matrix and Fiber respectively. The volumes and masses of these are

being computed with composite rule mixture computation.

$$\begin{bmatrix} V_f & V_m \end{bmatrix} = \begin{bmatrix} v_f & v_m \end{bmatrix} \begin{bmatrix} V_c & 0 \\ 0 & V_c \end{bmatrix} \quad (8)$$

The masses of polymer matrix and fiber vary proportionally with the product of their density and volume respectively.

$$\begin{bmatrix} M_f & M_m \end{bmatrix} = \begin{bmatrix} V_f & V_m \end{bmatrix} \begin{bmatrix} \rho_f & 0 \\ 0 & \rho_m \end{bmatrix} \quad (9)$$

The costs of the Polymer Matrix and fibers are computed with

$$\begin{bmatrix} C_f^1 & C_m^1 \end{bmatrix} = \begin{bmatrix} M_f & M_m \end{bmatrix} \begin{bmatrix} C_f & 0 \\ 0 & C_m \end{bmatrix} \quad (10)$$

The cost of composite material is sum of the cost of fibers and the polymer matrix.

##### Algorithm 1:

**Input** : Fiber reinforcement fiber data space F.

**Output**: A set of selected fibers class DF, each fiber length is greater  $15l_c$

##### Method:

**Step 1**: Create a node N.

**Step 2**: if samples are all of the same class, C then return N as leaf node labeled with the class C;

**Step 3**: If an attribute list is empty then return N as the leaf node labeled with class C.

[Select a test attribute and determine its class.]

**Step 4**: Select a test attribute, which is determined heuristically.

**Step 5**: For each f in F satisfying discrimination function  $d_i$ , at the  $i^{\text{th}}$  level create a child node for the parent node and labeled it

**Step 6**: Repeat through step 4 until conditional class = 0

##### Algorithm 2:

**Inputs**: Table  $F_{m \times n}(L_f, C_f, \rho_f, E_f)$  contains, length which is greater  $15l_c$ , cost, density and tensile modulus of fibers and elastic modulus of composite,  $E_c$ , fiber critical length  $l_c$ . Polymer Matrix table  $P_{m \times n}(C_m, \rho_m, E_m)$ .

**Output**: Least cost fiber associated to the given candidate Polymer Matrix.

##### Method

**Step 1**: obtain the output of classifier,  $F_{m \times n}(c, d, e)$  of fibers.

**Step 2**: compute Volume of composite,  $V_c$ .

**Step 3**: Compute for  $i=1..n$  do// for a Polymer Matrix and all the fib

##### Begin

$$v_f \leftarrow \frac{1}{(E_f - E_m)} \left( \frac{E_c}{\text{Cos } \theta} - E_m \right)$$

$$v_m \leftarrow \frac{1}{E_m} \left( \frac{E_c}{\cos\theta} - E_f v_f \right)$$

Compute  $[V_f \ V_m]$ ,  $[M_f \ M_m]$ ,  $[C_f^1 \ C_m^1]$ ,

Compute  $C_c = C_f^1 + C_m^1$

end

$$\text{Min}(C_c(i))_{i=1..n}$$

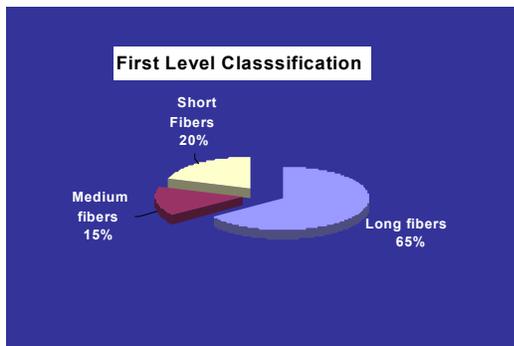
Step 4: Select

Step 5: stop

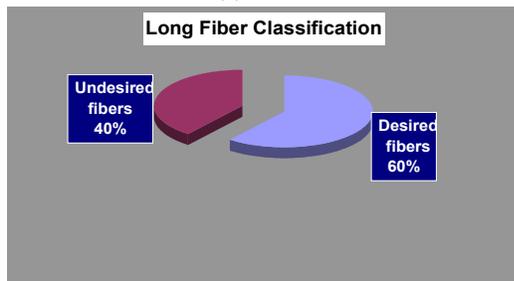
## 5. Results and Conclusion

### 5.1. Results

In this discussion of design requirements, a candidate polymer matrix Polycarbonate and a set of long fibers and their physical and mechanical properties motioned in section 2.1 are considered. The volume fractions, volumes, masses, costs of polymer matrix and long fibers and total cost of composite polymer are being computed. These are constrained to  $E_c = 120$  GPa, and  $V_c = 2222 \text{ cm}^3$  of polymer composite and critical length,  $l_c = 0.25 \text{ cm}$  of fibers. The decision tree classifier is tested with 250 reinforcement fibers. Percentages of reinforcement fibers classified by the liner discrimination function with selection quantitative parameter, w, equivalent to 1 at different levels of classes shown below.



(a)



(b)

Fig.3. (a) shows the first level classification and (b) shows the second level classification.

Since discrimination function is linear in nature, long fiber selection throughput is high as the w increases and is depicted as in fig.4.

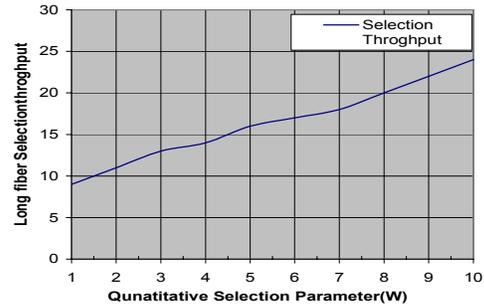


Fig.4 Linear function cost output.

The data base having 268 reinforcement fibers is considered as sample data set and fed it as input to decision tree classifier. The discrimination function being implemented with DT, classifies reinforcement fibers into long fiber class with 65%, medium class fibers with 15%, and short class fibers with 20%. In the second level, long fiber class is discriminated into desired fiber class 60% and undesired fiber class with 40%.

The volume fractions, Volumes, Masses, Costs of long reinforcement fibers of polymer matrix: polycarbonate and of the desired long fibers are computed. Knowledge about the least polymer composite is discovered from the database. Fig 5 shows top 9 reinforcement fibers whose modulus of elasticity greater than the modulus of elasticity of composite component. Total cost polymer composite is computed and heuristic method is implemented for the least cost selection. Fig.5 shows that the Aramid(Kevlar 49) among 9 fibers is the most suitable one for the given polymer matrix polycarbonate with least cost required for designing composite polymer for cylindrical pipe with longitudinal filament winding.

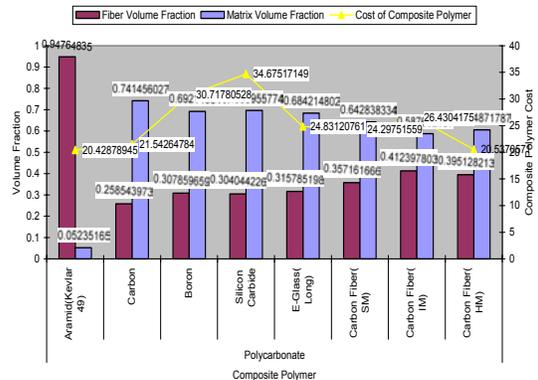


Fig.5. shows the volume fractions of both Polymer Matrix and Reinforcement fibers and the cost of

polymer composites: polycarbonate with different reinforcement fibers.

The followings are found by the analysis the output of the developed algorithms.

1. The volume fractions of Polymer Matrix and long fiber are such that  $v_f + v_m = 1$ .
2. The volume of a polymer composite depends on the volume both polymer matrix and reinforcement such that  $V_c = V_f + V_m$ .
3. A fiber reinforcement having high volume fraction leads more strength.

## 5.2. Conclusion

We have succeeded in developing knowledge discovery system for the selection of cost effective composite polymers for a cylindrical pipe with longitudinal filament winding through data mining approach. Two databases of Polymer matrix and Reinforcement fibers are considered for this system. For a given candidate polymer matrix in PM database, the system discovers a suitable long fiber in the RF database that meets the design requirements and for which the total cost of polymer composite is minimal. This system assists the design engineers in decision making on fiber selection for a given polymer matrix and derives hidden knowledge for composite polymers. In concurrent engineering applications, selection process is made easy for more production throughput.

## References

- [1]. V. Weiss. Computer-Aided Materials Selection, ASM Metal Handbook, Vol.20, Volume Chair George Dieter, ASM International Material Park, OH, 1997, pp309-314.
- [2]. P.A. Guteridge and J Tuner, "Computer Aided Material Selection", Materials and Design, 3(Aug).504-51(1982).
- [3]. L. Olsson, U. Bengtson, and H. Erichmeister, "Computer Aided Material Selection and Design in Computers in Material Technology.", T.Ericsson(cd.), Pergamon, Oxford, 1981, pp.17-25.
- [4]. P.P. Dargie, K. Parmeshwar, and W. R. D. Wilson, "MAPS 1: Computer Aided Design Systems for Preliminary Material and Manufacturing Process Selection.", Trans. ASME, Journal of Mech.Design.104, 126-136(1982).
- [5]. D.Cebon and M F Ashby "The Optimal Selection of Engineering Entities", Technical report CUED/C-EDC/TR 59, Nov.1997, ISSN 0963-5432.
- [6]. <http://www.matweb.com/>
- [7]. D. Cebon and M F Ashby "Materials Selection For Precision Instruments", Materials Science and Technology. Vol.5, pp 26-306, 1994.
- [8]. M F Ashby "Selecting the Components of Composites", Journal of the physique III, Vol.3, November 1993.
- [9]. M F Ashby "Material Selection Charts", ASM Metals Handbook, Vol.20, Volume chair George Dieter, ASM International, Materials Park, OH, 1997, pp266-280.
- [10]. D. Cebon and M F Ashby and "Computer-Based Materials Selection for Mechanical Design" in Computerized and Networking of material Databases ", ASTM STP 1140, Philadelphia, 1992.
- [11]. UlrikeG K, Wegst and Michae F Ashby "The Development And Use Of A Methodology For The Environmentally-Concious Selection Of Materials", Proceedings of the international conference of Integrated design and Process Technology (IDPT), Vol. 5, pp88-93.
- [12]. Bergamaschi, S Bombarda, G Piancastelli, L Sartori C," An Expert System for The Selection of a Composite Material", Second international conference on Engineering", Vol. 2, pp 140-141.
- [13]. D.Cebon and M F Ashby "The Optimal Selection of Engineering Entities", Technical report CUED/C-EDC/TR 59, Nov.1997, ISSN 0963-5432.
- [14]. Crestine Gertosta and Alan Dussauchoy," Knowledge Discovery from Industrial Databases", Journal of Intelligent Manufacturing, 15, 29-37, 2004.
- [15]. S. Rasoul Safavina and David Landgreb,"A Survey on Decision Tree Classifier Methodology", IEEE Transaction on Systems, Man and Cybernetics, Vol.21, No.3, pp 660-674, May 1991.
- [16]. William D.Callister, Jr., "Materials Science and Engineering An Introduction", John Wiley & Sons, Inc, 1997.