Proceedings of the World Congress on Engineering 2007 Vol I
WCE 2007, July 2 - 4, 2007, London, U.K.

# A Fuzzy Entropy Algorithm For Data Extrapolation In Multi-Compressor System

Gursewak S Brar[#], Yadwinder S Brar[$], Yaduvir Singh[*]

*Abstract*-- **In this paper incomplete quantitative data has been dealt by using the concept of fuzzy entropy. Fuzzy entropy has been used to extrapolate the data pertaining to the compressor current. Certain attributes related to the compressor current have been considered. Test data of compressor current used in this knowledge discovery algorithm knows the entire attribute clearly. The developed algorithm is very effective and can be used in the various application related to knowledge discovery and machine learning. The developed knowledge discovery algorithm using fuzzy entropy has been tested on a multi-compressor system for incomplete compressor current data and it is found that the error level is merely $\pm 4.40\%$, which is far better than other available knowledge discovery algorithms.**

*Key words*: **fuzzy entropy, genetic programming, incomplete data, classification, and knowledge discovery, multi-compressor system**

## I. INTRODUCTION

Multi-compressor systems are used for refrigeration, cooling and air-conditioning. These systems are quite different as compared to electrical or electronic systems because of these natural attributes like weight, inertia, force and torque requirements etc. As reported in literature there is trend to use computerized numerical techniques, which greatly reduce energy, time, cost etc and drastically enhances the efficiency of multi-compressor systems. There is trend to design controllers for mechanical systems based on data instead of models. The data is normally huge in size and vary in nature. Data base controller depends upon selection and use of right kind of data. Methods of data interpolation and data extrapolation are applied in a great variety of areas for data mining and knowledge discovery, forecasting and systems modeling, optimization and pattern recognition. Inductive MDH algorithms give possibility to find automatically interrelations in data, to select optimal structure of model or network and to increase the accuracy of existing algorithms. The algorithms based on fuzzy entropy will be used to design a data based controller for the system. [4]

In recent years machine learning and knowledge discovery techniques have attracted a great deal of attention in the information area. Classification is one of the important research topics on these research areas.

[#]Department of Electrical Engineering
Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib, Punjab
[$]Department of Electrical Engineering
Giani Zail Singh College of Engineering and Technology, Bhatinda, Punjab
[*]Department of Electrical and Instrumentation Engineering
Thapar University, Patiala , Punjab
Email: brargs77@rediffmail.com

Most of researches on classification concern that a complete data set is given as a training set and the test data know all values of attributes clearly. Unfortunately, incomplete data are commonly seen in real world applications. Knowledge discovery algorithms take an input of training examples of target knowledge, and output a fuzzy logic formula that best fits the training examples. The execution is done in some steps and it could be made possible by using object, data input, algorithm, process, experiment, and results. [1], [2], [10]

Fuzzy Logic is a form of logic that extends on Boolean logic that incorporates partial values of truth - Instead of sentences being "Completely true" or "Completely false," they are assigned a value that represents their degree of truth. In fuzzy systems, values are indicated by a number (called a truth value) in the range from 0 to 1, where 0.0 represents absolute falseness and 1.0 represents absolute truth. Fuzzification is the generalization of any theory from discrete to continuous. [9] Fuzzy Logic is important to AI because they allow computers to answer 'to a certain degree' as opposed to in one extreme or the other. In this sense, computers are allowed to think more 'human-like' since almost nothing in our perception is extreme, but is true only to a certain degree. Through fuzzy logic, machines can think in degrees, solve problems when there is no simple mathematical model, solve problems for highly nonlinear processes and use expert knowledge to make decisions. [6]

Knowledge discovery in Fuzzy logic is based on membership function values. After a fundamental algorithm, fuzzy logic functions are applied to a more practical example of classification problem, in which expressiveness of fuzzy logic functions is examined for a well-known machine-learning database. Here in this work, we have investigated the problem of incomplete data in data sets in the input-output behavior of a multi-compressor system. A data set with at least one missing attribute value is referred as an incomplete data set. Since the incomplete samples don't provide perfect information for training process, most of the traditional classification algorithms cannot be with incomplete data directly but generate in accurate classifiers from an incomplete data. Hence the incomplete data must be tackled well so that good classification models can be developed for real life applications. The genetic programming is one of the techniques on evolutionary computation. The genetic programming has been applied to several applications like symbolic regression, the robot control programs, and classification, etc. genetic programming can discover underlying data relationships and present these relationships by expression. A supervised learning method based on genetic programming to handle the classification problem with incomplete data in attributes has been used. [3],[5],[7],[8]

In this paper a new strategy based on Fuzzy Entropy has been introduced for the first time to deal with the incomplete quantitative data in the case of multi-compressor system. For handling incomplete quantitative data, we have firstly applied fuzzy entropy to discriminate the best number of intervals, which have been granulated as a fuzzy linguistic term with a membership function. Then, we employ the linguistic term to infer the missing attribute values based on the max-min composition method according to their class labels. This paper also introduces a supervised learning method based on genetic programming to handle the classification problem with incomplete data in attributes.

## II. CASE STUDY: MULTI-COMPRESSOR SYSTEM

This plant is basically a chemical plant and used for making the food products. In this temperature variations occur, so cooling is required from time to time. A robust controller is required, which can provide temperature stabilization and accurate cooling.

The important units of the system are engine, refrigerator and heat pump. These are shown below in figure 1. Systems having thermodynamic importance are divided into two groups. First, work developing systems which includes all types of engines producing power using thermal energy and second work-absorbing systems which include compressors, refrigerators and heat pumps etc. Source and sink contain infinite energy at constant temperature. Source temperature is always higher than the sink temperature.

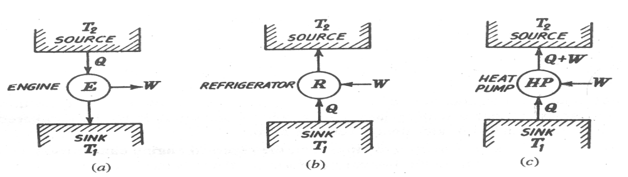


Figure 1: Engine, refrigerator and heat pump

In case of the engine, for higher efficiency it is desired to get maximum amount of work W, with minimum supply of energy Q. The performance of an engine is taken into account by the ratio W/Q, which is known as efficiency (η) of the engine and is given as below.

$$\eta = W/Q \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

In case of refrigerator, it is desired to maintain temperature $T_1<T_2$, where $T_2$ is the atmospheric temperature. For greater economy, the maximum Q must be taken from sink with the minimum amount of W, so that the performance of the refrigerator is taken into account by a ratio Q/W. The theoretical coefficient of performance (C.O.P.) is calculated as below.

$$\text{C.O.P.} = Q/W \ldots\ldots\ldots\ldots\ldots (2)$$

Also

$$\text{Relative C.O.P.} = (\text{actual C.O.P./theoretical C.O.P}) \ldots\ldots (3)$$

The cycle used for refrigerator is also used for heat pump. The performance of the heat pump is taken into account by a ratio (Q+W)/W and it is known as energy performance ratio (E.P.R.) It is obtained as below.

$$\text{E.P.R.} = (1 + Q/W) \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Also

$$\text{E.P.R.} = (\text{C.O.P.} + 1) \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

The value of C.O.P. should be less than one or greater than one, which depends upon the type of the refrigeration system. The value of E.R.P. should always be greater than one. Figure 2 shows the multi-mode system with single compressor, which is used when numbers of loads at same temperatures are to be taken by the refrigerating plant.

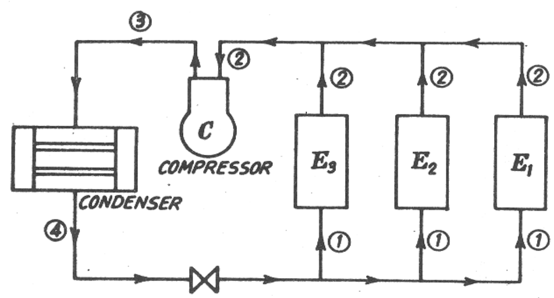


Figure 2: Multimode systems with single compressor

The arrangement of multi-evaporators at different temperatures with back pressure valves is shown in figure 3.and 1 is the condition of the refrigerant entering into the evaporator $E_1$ and leaving with condition 2.Then 8 is the condition of the refrigerant entering into the evaporator $E_2$ and leaving with condition 3.Then 7 is the condition of the refrigerant entering into the evaporator $E_3$ and leaving with condition 4.

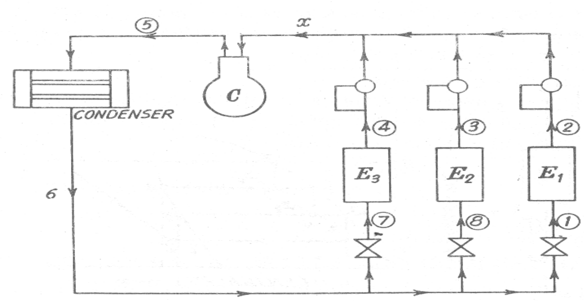


Figure 3: Multi-evaporators at different temperatures with backpressure valves

The pressures of the refrigerants coming out of the evaporators and after leaving the back pressure valves is same and that is the suction pressure of the compressor. Table 1 as given below, provides the operational data taken over a particular period as a sample.

Table 1: Observation data for a single compressor in a multi-compressor system

| S. No. | Suction pressure Kg/cm$^2$ | Oil pressure Kg/cm$^2$ | Delivery pressure Kg/cm$^2$ | Current Amperes |
|---|---|---|---|---|
| 1 | 3.6 | 5.6 | 10.2 | 120 |
| 2 | 3.7 | 5.7 | 11.2 | 128 |
| 3 | 3.8 | 6.0 | 11.6 | 133 |
| 4 | 3.6 | 5.6 | 12.4 | 136 |
| 5 | 3.5 | 5.6 | 12.0 | 142 |
| 6 | 2.9 | 4.26 | 11.6 | 156 |
| 7 | 2.8 | 4.25 | 10.8 | 176 |
| 8 | 2.6 | 5.8 | 10.4 | 190 |
| 9 | 2.4 | 5.9 | 9.9 | 200 |
| 10. | 2.3 | 6.0 | 9.3 | 220 |

While Suction pressure of the compressor-system decreases, current taken increases monotonically. However the two pressures viz. Oil pressure and Delivery pressure exhibit swing behavior.

## III. FUZZY ENTROPY

It is a measure of the amount of uncertainty of fuzzy set. Fuzzy entropy discriminates the best number of intervals for the quantitative attribute. The fuzzy entropy for each interval is defined as below. [9]

1. Let X = {$x_1$, $x_2$ …$x_n$} be a universal set with elements xi distributed in a pattern space, where I = 1, 2…n.
2. Let A be a fuzzy set defined on an interval of pattern space, which contains k element (k<n). The mapped membership degree of the element $x_i$ with the fuzzy set A is denoted by μA ($x_i$).
3. Let $C_1$, $C_2$ …$C_k$ represent k classes into which the n elements are divided.
4. Let Sf ($x_n$) denotes a set of element of class j on the universal set X. it is a subset of universal set X.
5. The match degree Dj with the fuzzy set A for the elements of class j in an interval, where j = 1,2………..k is defined as

$$Dj = \sum_{X\in Sj\,(xn)} \mu A(r) \Big/ \sum_{x\in S} \mu A(r) \quad \ldots\ldots\ldots\ldots\ldots (1)$$

6. The fuzzy entropy FEj (A) = - Dj log2 Dj
7. The fuzzy entropy FE (A) on the universal set x for the element within an interval is defined as

$$FE (A) = \sum_{j=1}^{k} FECj (A). \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

## IV. PROBLEM FORMULATION, MATHEMATICAL MODELING AND ALGORITHM DEVELOPMENT

We have used symmetrical triangular form of membership function to represent fuzzy set for the complete data as shown in table 1.The optimum value on the basis of above membership function values is obtained as below.

Optimum value = ∑ If*μ / ∑ μ ………………(3)

The developed algorithm for correct membership value on the basis of Fuzzy Entropy in this case, is given below.

Step 1: Input: An incomplete quantitative data set in which the number of data and those values are given.

Step 2: Output: Assign a membership value function and find optimum value among those.

```
Step 3:  Nd = number of data given.
    V [I] =values of given data.
    Rmax = Maximum value of given dataset.
      Rmin = Minimum value of given dataset.
    T = number of triangle between Rmin and Rmax.
        Initialize
                                        T = 2
        Value of Fuzzy Entropy, FE (previous) = 0
        Value of Fuzzy Entropy, FE (present)  = 0
        Value of optimum data, Vopt (previous) = 0
        Value of optimum data, Vopt (present) = 0
Step 4:
Do
  {
   FE (previous) = FE (present)
   FE (present)   = 0
   Initialize the match degree D [j] = 0
              Numerator Nr [j] = 0
              Denominator Dr = 0
   For (I =0; I<= Nd; I++)
    {
    For (J=0; J<= T; J++)
  {
   If (Rmin+ (Rmax*(J-1)/ (T+1)) < V [I] <= Rmin+
(Rmax*(J-1)/ (T+1)))
    {
   Nr [J] =Nr [J]-(V [I]*(T+1)/ (Rmax-Rmin)) +
((T+1)*(Rmin+ (Rmax-Rmin)*(J+1)/ (T+1))/ (Rmax-Rmin));
        Nr [j+1] =Nr [j+1] + (V [I]*(T+1)/ (Rmax-Rmin))-
((T+1)*(Rmin+ (Rmax-Rmin)*J/ (T+1))/ (Rmax-Rmin));
    }
   }
    For (J=0; J<= 2.T; J=J+2)
  {
   If ((Rmin+ (Rmax-Rmin)*J/ ((T+1)*2)) <V [I] <= (Rmin+
(Rmax-Rmin)*(J+1)/ ((T+1)*2)))
        {
      Msp=-(V [I]*(T+1)/ (Rmax-Rmin)) + ((T+1)*(Rmin+
(Rmax-Rmin)*(J+2)/ ((T+1)*2))/ (Rmax-Rmin));
         Nropt=Nropt + Msp*V [I];
      }
```

```
        If ((Rmin+(Rmax-
Rmin)*(j+1)/((nt+1)*2))<V[I]<=(Rmin+(Rmax-
Rmin)*(j+2)/((nt+1)*2)))
          {
        Msp= (V [I]*(T+1)/ (Rmax-Rmin))-((T+1)*(Rmin+
(Rmax-Rmin)*(j+2)/ ((T+1)*2))/ (Rmax-Rmin));
          Nropt=Nropt + Msp*V [I];
          }
             Dr=Dr+Msp;
  }
     }

     For (J=1; J<=T; J++)
  {
   D [J] = Nr [J] / Dr
   FE (present) = FE (present) – D [J]. log2 D[J]
  }

 Vopt (previous) = Vopt (present)
 Vopt (present) = Nropt/Dr
 T++
   } While (FE (present) > FE (previous))

 Optimum value of given dataset = Vopt (previous)
```

Step5: T = T-2

T is number of symmetrical triangle exist between $R_{min}$ and $R_{max}$. Use this Fuzzy membership function value for finding missing values. Defuzzify the function for finding the optimum value.

Figure 4, 5, 6 and 7 depict fuzzy sets showing four data classifications. Triangular fuzzy sets have been chosen to keep the mathematical model linear.

In complex nonlinear knowledge discovery problems some suitable nonlinear fuzzy sets like sigmoidal, trapezoidal etc can be chosen.

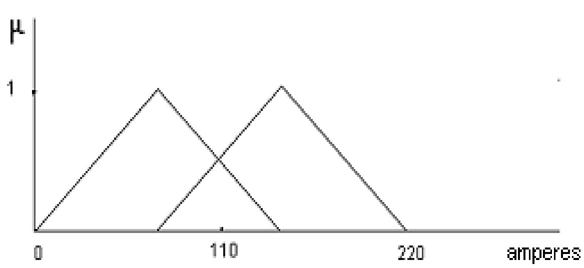


Figure 4: Data classification "1" for assigning membership values

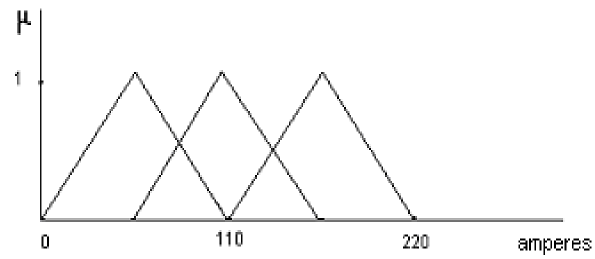


Figure 5: Data classification "2" for assigning membership values

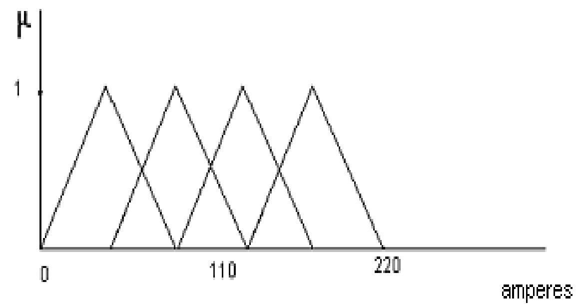


Figure 6: Data classification "3" for assigning membership values

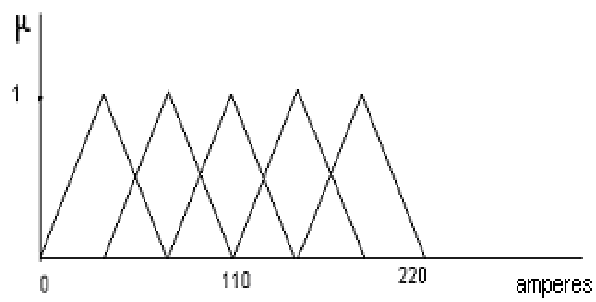


Figure 7: Data classification "4" for assigning membership values

Here Fuzzy Entropy (FE) is minimum, when number of class interval is 2. So we will chose membership value according to this interval.

ISBN:978-988-98671-5-7 

Table 2: Fuzzy Entropy (FE) for desired optimum value of compressor output pressure

| Sr.No. | No of class between 0 and max.input (based on membership function) | Total Fuzzy Entropy (FE) |
|---|---|---|
| 1 | 2 | 0.869630672 |
| 2 | 3 | 1.424813196 |
| 3 | 4 | 2.017413200 |
| 4 | 5 | 2.159269261 |

Now putting the values in Eq. 3, we get

Optimum value = {(123*2.7 + 151*3.7 + 156*3.5 + 156*3.5 + 165*3 + 172*2.6 + 176*2.4 + 190*1.6 + 200*1.1 + 220*0)/4}/{(2.7+3.7+3.5+3.5+3+2.6+2.4+1.6+1.1+0)/4}
= 160.64 amperes

From above, the optimum value for the compressor current is equal to 160.64 amperes.

## V. SIMULATION AND TESTING

The above algorithm for Fuzzy Entropy has been coded in a higher-level language. A set of 10 incomplete data pertaining to the compressor current in a multi-compressor system have been fed to the code and the simulation results have been obtained. The testing results are given in table 3 below.

Table 3: For calculating the accuracy of this technique

| Sr. No. | Data range | Desired data value of compressor current ID (average value of input data's) | Computed value compressor current from this algorithm IC | Error = $\frac{(IC- ID)100}{IC}$ (In %) |
|---|---|---|---|---|
| 1 | 123,151,156,156,165,172176,190,200,220 | 170.9 | 160.64 | +6.01 |
| 2 | 127,135,158,170,150,186,210,202,166,156 | 166.0 | 158.03 | +4.8 |
| 3 | 148,162,160,188,200,195,190,172,170,180 | 176.5 | 170.85 | +3.2 |
| 4 | 155,160,162,190,176,152,158,205,180,150 | 168.8 | 160.02 | +5.2 |
| 5 | 140,148,173,200,210,205,164,157,151,180 | 172.8 | 163.98 | +5.1 |
| 6 | 154,158,191,176,200,168,165,158,195,188 | 175.3 | 182.48 | -4.1 |
| 7 | 160,178,192,144,146,198,200,151,158,149 | 167.6 | 157.37 | +6.1 |
| 8 | 168,172,165,161,183,195,193,180,175,200 | 179.2 | 184.39 | -2.9 |
| 9 | 155,159,163,168,185,191,186,205,207,196 | 181.5 | 187.85 | -3.5 |
| 10 | 168,153,156,178,182,191,202,195,180,175 | 178.0 | 172.48 | +3.1 |

Average error in compressor current = $\pm$4.40%

Figure 8 provides a measure of the compressor current fuzzy entropy (error) in the case of data extrapolation in the case of multi-compressor system. When we increases the number of class interval between maximum and minimum range of given datasets, then overlapping area will increase. According to Fuzzy Entropy (FE) formula, always it will be in –sign for each interval, so as we increases the number of class interval Fuzzy Entropy will increases in –ve direction.

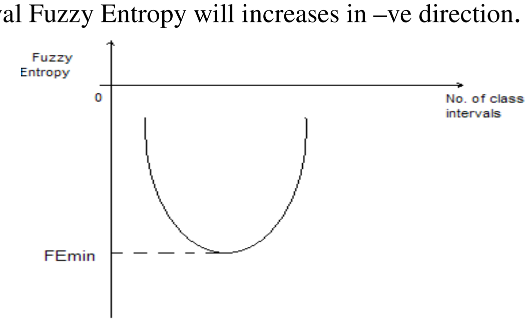


Figure 8: Fuzzy entropy as a measure of error

After a particular value of class interval, FE will start to decrease again. We can detect easily the number of class interval when the FE is minimum and can say that it is the best suitable classification for those given datasets on the basis of minimum Fuzzy Entropy.

## VI. RESULTS AND DISCUSSIONS

The results as given in table 3 have been checked with their actual values. The average error in compressor current for the data set comprising of ten uncertain values for this case study comes out to be $\pm$4.40%. Here we are observing that computed value is deviating very less from observed value, but at this computed value fuzzy entropy is minimum so this technique gives the better option to select the value for further processing.

## VII. CONCLUSIONS

Fuzzy entropy is one of the best knowledge discovery methods as soon in the paper data classification is one of the important research areas. Fuzzy entropy is very efficient in handling incomplete qualitative and nominal data. The results of the proposed algorithm demonstrate an error of $\pm 4.40\%$. The knowledge discovery based on fuzzy entropy has been made further efficient by using a supervised by learning based on genetic programming. Genetic programming is one of the best available methods for generating the data classifiers.

## REFERENCES

[1] J.W. Grzymala-Busse and M. Hu, "A comparison of several approaches to missing attribute value in data mining," in Proc. Of Second International conference on rough Sets and current trends In Computing, RSCTC2000, pp.378-385, 1997.

[2] C. Black, E. Keogh and C.J. Merz, "USI repository of machine learning database," Irvine, University of California, Department of Information and computer science (1998).

[3] J.H. Friedman, "A recursive partitioning decision rule for non-parametric classification," IEEE trans. On computer science, 1999,pp.404-408.

[4] Zhou W, Kim, J. Soedel W., "New iterative scheme in computer simulation of positive displacement compressor considering the effect of gas pulsation", Transactions of the AMSE 123 pp 282-288 2001.

[5] R. Slowinski and J.Stefanowski, "Handling various types of uncertainty in the rough set approach," in Proc. Of the international workshop on Rough sets and Knowledge Discovery, 1993,pp.366-376.

[6] T-P Hong, L-H Tseng and B-C. Chien, "Learning Fuzzy rules from incomplete quantitative data by rough sets", in Proc. Of the 2002 IEEE International Conference on Fuzzy System, pp.1438-1443.

[7] J.K. Kishore, L.M. patnaik, V.K. agrawal, "Application of genetic Programming for Multicategory Pattern Classification," IEEE Trans. On evolutionary computation, vol.4, No. 3,2000,pp.242-258.

[8] M. Bramrier and W. Banzhaf, "A comparison of linear genetic programming and neural networks in medical data mining," IEEE trans. On Evolutionary Computation, Vol.5, No. 1 Feb 2001, pp. 17-26.

[9] B-C chien, J-h Yang, W-Y. Lin, "Generating effective classifiers with Supervised Learning of Genetic Programming," in Proc. Of the 5th international Conference on data warehousing and knowledge discovery, 2003, pp. 192-201.

[10] H. -M. Lee, C-M. Chen, and Y. -L. Jou, " An efficient fuzzy classifier with feature selection based on fuzzy entropy," IEEE trans. On System, Man, and Cybernetics-partB: cybernetics, Vol.31, No.3, 2001, pp.426-432.