A Feature Selection Method Based on ∩ - Fuzzy Similarity Measures Using Multi Objective Genetic Algorithm

Hassan Nosrati Nahook, Mahdi Eftekhari

Abstract — Feature selection (FS) is considered to be an important preprocessing step in machine learning and pattern recognition, and feature evaluation is the key issue for constructing a feature selection algorithm. Feature selection process can also reduce noise and this way enhance the classification accuracy. In this article, feature selection method based on ∩ - fuzzy similarity measures by multi objective genetic algorithm (FSFSM – MOGA) is introduced and performance of the proposed method on published data sets from UCI was evaluated. The results show the efficiency of the method is compared with the conventional version. When this method multi-objective genetic algorithms and fuzzy similarity measures used in CFS method can improve it.

Keywords: Feature Selection, Fuzzy Similarity Measures, Multi Objective Genetic.

I. INTRODUCTION

Large dimensionality presents a problem for handling data due to the fact that the complexity of many commonly used operations are highly dependent (e.g. exponentially) on the level of dimensionality. The problems associated with such large dimensionality mean that any attempt to use machine learning or data-mining tools to extract knowledge, results in very poor performance. Feature selection (FS) [1] is a process which attempts to select features which are information-rich whilst retaining the original meaning of the features following reduction. Most learning algorithms are unable to consider problems of such size, whilst those that are not will usually perform poorly.

A similarity measure is an important tool for determining the degree of similarity between two objects. Kaufman and Rousseeuw [2] presented some examples to illustrate traditional similarity measure applications in hierarchical cluster analysis. Since Zadeh [3] originated the idea of fuzzy sets, many different similarity measures between fuzzy sets have been proposed in the literature. Zwick etal [4]. Reviewed geometric distance and Hausdorff metrics presenting similarity measures among fuzzy sets. Pappis and Karacapilidis [5] proposed three similarity measures based on union and intersection operations, the maximum difference, and the difference and sum of membership grades. Wang [6] presented two similarity measures between fuzzy sets and between elements. Liu [7] and Fan and Xie [8] provided the axiom definition and properties of similarity measures between fuzzy sets.

Hassan Nosrati Nahook, MSc Computers - AI, Computer Engineering Department, Science and Research Branch, Islamic Azad University, Kerman, Iran .

Mahdi Eftekhari, Assistant Professor, Department Computer Engineering - Faculty of Engineering - Shahid Bahonar University of Kerman, Iran.

Turksen and Zhong [9] applied similarity measures between fuzzy sets for approximate analogical reasoning. Buckley and Hayashi [10] used a similarity measure between fuzzy sets to determine whether a rule should be fired for rule matching in fuzzy control and neural networks.

The theory of fuzzy sets, proposed by Zadeh [3], has gained successful applications in various fields. Measures of similarity between fuzzy sets, as an important content in fuzzy mathematics, have gained attention from researchers for their wide applications in real world. Based on similarity measures that are very useful in some areas, such as pattern recognition, machine learning, decision making and market prediction, many measures of similarity between fuzzy sets have been proposed and researched in recent years.

In this paper, we will review several popular similarity measures between fuzzy sets and a feature selection method based on fuzzy similarity measures by multi objective genetic algorithm (FSFSM – MOGA) is presented. This paper is organized as follows. Section 2 describes the related terms. Section 3 explains the multi – objective genetic algorithm. Section 4 describes the correlation based on feature selection method. Section 5 explains the proposed method based on fuzzy similarity measure for evaluate similarity between features by multi objective genetic algorithm. In Section 6, the experimental results of the proposed method are presented. The last section summarises and conclusion related work.

II. RELATED TERMS

A. T – norm and T – conorm

The triangular norms (t-norm), which generalize the form of intersection and union, are next well described and later will be used to construct our similarity measure:

For any x, y, z and $u \in [0, 1]$.

T – norm: A two-place function $T:[0, 1] \times [0,1] \rightarrow [0,1]$ is called t – norm if the following conditions are satisfied:

1. $T(x,1) = x$: one identity;
2. $x \leq z, \ y \leq u \Rightarrow T(x, y) \leq T(z, u)$: monotonicity;
3. $T(x, y) = T(y, x)$: commutivity;
4. $T(T(x, y), z) = T(x, T(y, z))$: associativity;

A t-norm is called Archimedean if and only if T is continuous and $\forall x \in [0, 1]: T(x, x) < x$.

T – conorm: A two-place function $Sn:[0, 1] \times [0,1] \rightarrow [0,1]$ is called t – conorm if the following conditions are satisfied:

1. $Sn(x,0) = x$: zero identity;
2. $x \leq z, \ y \leq u \Rightarrow Sn(x, y) \leq Sn(z, u)$: monotonicity;
3. $Sn(x, y) = Sn(y, x)$: commutivity;
4. $Sn(Sn(x, y), z) = Sn(x, Sn(y, z))$: associativity;

Notice that t-norms are functions which are called fuzzy intersections and unions are the common shorthand term for

triangular norms, t-norm and t-conorm only differ on their boundary conditions. Some additional properties of t-norm and t-conorm are presented in the following definitions [11].

A function $Sn:[0, 1]\times[0,1]\to[0,1]$ is dual t-conorm of t-norm such that for all $x, y\in[0, 1]$ both the following equivalent equalities hold, $Sn(x, y)=1-T(1-x,1-y)$ and $T(x, y)=1-Sn(1-x,1-y)$, where $(1 - x)$ and $(1 - y)$ are respectively complements of x and y.

Next we present a list of the main well know and most frequently used t – norms [11], [12]:

$$T_{min}(x, y) = \min(x, y): \text{Minimum;} \qquad (1)$$

$$T_{prod}(x, y) = xy: \text{Algebraic product;} \qquad (2)$$

$$T_{bprod}(x, y) = \max(0, x+y-1): \text{Bounded product;} \qquad (3)$$

$$T_{Ds}(x, y) = \begin{cases} x & ;y=0 \\ y & ;x=0 \\ 1 & \text{otherwise} \end{cases} : \text{Drastic sum;} \qquad (4)$$

$$T_{\alpha}^{H}(x, y) = \frac{xy}{\alpha+(1-\alpha)x + y - xy}, \alpha\geq 0: \text{Hamacher's t-norm ;} \qquad (5)$$

$$T_{\beta}^{F}(x, y) = \log_{\beta}\left(1+\frac{(\beta^{x}-1)(\beta^{y}-1)}{\beta-1}\right), \beta > 0, \beta \neq 1: \text{Frank's t-norm;} \qquad (6)$$

$$T_{\gamma}^{Y}(x, y) = 1-\min(1, (1-x)^{\gamma}+(1-y)^{\gamma})^{\frac{1}{\gamma}}, \gamma>0: \text{Yager's t-norm;} \qquad (7)$$

$$T_{k}^{D}(x, y) = 1-\frac{1}{1+\left(\left(\frac{1-x}{x}\right)^{k}+\left(\frac{1-y}{y}\right)^{k}\right)^{\frac{1}{k}}}, k>0: \text{Dombi's t-norm ;} \qquad (8)$$

$$T_{\theta}^{W}(x, y) = \max\left(0, \frac{x+y-1+\theta xy}{1+\theta}\right), \theta>-1: \text{Weber's t-norm ;} \qquad (9)$$

$$T_{SS}^{l}(x, y) = \max\left(0, (x^{p}+y^{p}-1)^{\frac{1}{p}}\right), p>0: \text{Schweizer Sklar's t-norm;} \qquad (10)$$

$$T_{\lambda}^{Yu}(x, y) = \max(0, (1+\lambda)(x+y-1)-\lambda xy), \lambda>-1: \text{Yu's t-norm;} \qquad (11)$$

By using the duality we can easily establish the Yu's t-conorm, which is:

$$Sn_{\lambda}^{Yu}(x, y) = \min(1,x+y+\lambda xy), \lambda>-1 \qquad (12)$$

B. Similarity measures for Fuzzy sets

In this section we present a brief review of similarity measures for fuzzy sets and their axiomatic basis. Since the concept of similarity has a wide range of applications, there are different approaches present in literature as axioms for degree or measure of similarity. These axioms have differences and similarities depending upon the contexts in which they are constructed. At first hand, a similarity measure for fuzzy sets is expected to be a *T*- equivalence on $F(X)$, which is later realized to be a very unrealistic requirement. Some other lists of properties are also found in literature that a reasonable similarity measure must satisfy. We shall suffice to present a set of axioms formulated by Bustince [13] for an interval valued similarity measure.

A function $\varsigma:F(X)\times F(X)\to[0, 1]$ is called a normal interval valued similarity measure, if ς satisfies following properties for all A, B, C $\in F(X)$:

I. $\varsigma(A, B)=\varsigma(B, A)$,

II. $\varsigma(A, A^{c})=[0, 0]$,

III. $\varsigma(A, A)=[1, 1]$,

IV. Monotonic
if $A\subseteq B\subseteq C$, then $\varsigma(A, B)\geq\varsigma(A, C)$ and $\varsigma(B, C)\geq\varsigma(A, C)$

C. Fuzzy similarity measures

(1) Simple fuzzy similarity measures

As definition of the cardinality of a fuzzy set A in X we consider the usual sigma-count of A:

$$\#A=\sum_{i=1}^{n} A(x_i) \qquad (13)$$

Furthermore, the complement A^c of A is defined by:

$$A^c(x_i)=1-A(x_i) \qquad (14)$$

and therefore $\#A^c = n - \#A$.

We have expressed T – norms in Section 2.1. In this paper, only the T – norm of equation (1) we use. Consider two fuzzy sets A and B in X and let $a_i = A(x_i)$ and $b_i = B(x_i)$, then we define:

$$A \cap B(x_i) = T(a_i,b_i) \qquad (15)$$
$$A\cup B(x_i)=Sn(a_i, b_i) \qquad (16)$$

Where T is an arbitrary t-norm, and Sn denotes its dual t-conorm: $Sn(x, y) = 1-T(1-x,1-y)$. We further restrict the t-norm T to the family of Minimum t-norms, namely the t-norms characterized by the functional equation:

$$Sn(x, y)+T(x,y)=x+y \qquad (17)$$

Hence, fuzzification equation (16) for set union can be restated in the alternative form:

$$A\cup B(x_i)=a_i+b_i-T(a_i, b_i) \qquad (18)$$

Notice that rules (13), (14) and (18) are such that both the expressions $\#(A\cup B)+\#(A\cap B)$ and $\#A + \#B$ are fuzzified to the same expression.

Equations (13) and (14) leads to the fuzzy similarity measures listed in Table (1).

Table (1): Simple fuzzy similarity measures.

S	Expression
S_{11}	$\dfrac{\min\left(\sum a_i, \sum b_i\right)}{\max\left(\sum a_i, \sum b_i\right)}$
(Complement S_{11})S_{12}	$\dfrac{n - \max\left(\sum a_i, \sum b_i\right)}{n - \min\left(\sum a_i, \sum b_i\right)}$
S_{13}	$\dfrac{\min\left(\sum a_i, \sum b_i\right)}{n}$
(Complement S_{13})$S1_4$	$\dfrac{n - \max\left(\sum a_i, \sum b_i\right)}{n}$

(2) ∩ - based fuzzy similarity measures

From Table (1) we see that are candidate for fuzzification by means of equations ((13), (14) and (18). In Table (2) are shown the expressions of the corresponding fuzzy similarity measures [14].

Table (2): ∩ - based fuzzy similarity measures.

S	Expression
S_{21}	$\dfrac{\sum T(a_i,b_i)}{\max\left(\sum a_i,\sum b_i\right)}$
(Complement S_{21}) S_{22}	$\dfrac{n-\sum a_i-\sum b_i+\sum T(a_i,b_i)}{n-\min\left(\sum a_i,\sum b_i\right)}$
S_{23}	$\dfrac{\sum T(a_i,b_i)}{\sum a_i+\sum b_i-\sum T(a_i,b_i)}$
(Complement S_{23}) S_{24}	$\dfrac{n-\sum a_i-\sum b_i+\sum T(a_i,b_i)}{n-\sum T(a_i,b_i)}$
S_{25}	$\dfrac{\min\left(\sum a_i,\sum b_i\right)}{\sum a_i+\sum b_i-\sum T(a_i,b_i)}$
(Complement S_{25}) S_{26}	$\dfrac{n-\max\left(\sum a_i,\sum b_i\right)}{n-\sum T(a_i,b_i)}$
S_{27}	$\dfrac{\sum T(a_i,b_i)}{n}$
(Complement S_{27}) S_{28}	$\dfrac{n-\sum a_i-\sum b_i+\sum T(a_i,b_i)}{n}$

III. MULTI – OBJECTIVE GA

In GA terminology, a solution vector $x \in X$ is called an individual or a chromosome. Chromosomes are made of discrete units called genes. Each gene controls one or more features of the chromosome. In the original implementation of GA by Holland, genes are assumed to be binary digits. In later implementations, more varied gene types have been introduced. Normally, a chromosome corresponds to a unique solution x in the solution space. This requires a mapping mechanism between the solution space and the chromosomes. This mapping is called an encoding. In fact, GA work on the encoding of a problem, not on the problem itself.

GA operate with a collection of chromosomes, called a population. The population is normally randomly initialized. As the search evolves, the population includes fitter and fitter solutions, and eventually it converges, meaning that it is dominated by a single solution. Holland also presented a proof of convergence (the schema theorem) to the global optimum where chromosomes are binary vectors.

GA use two operators to generate new solutions from existing ones: crossover and mutation. The crossover operator is the most important operator of GA. In crossover, generally two chromosomes, called parents, are combined together to form new chromosomes, called offspring. The parents are selected among existing chromosomes in the population with preference towards fitness so that offspring is expected to inherit good genes which make the parents fitter. By iteratively applying the crossover operator, genes of good chromosomes are expected to appear more frequently in the population, eventually leading to convergence to an overall good solution.

The mutation operator introduces random changes into characteristics of chromosomes. Mutation is generally applied at the gene level. In typical GA implementations, the mutation rate (probability of changing the properties of a gene) is very small and depends on the length of the chromosome. Therefore, the new chromosome produced by mutation will not be very different from the original one. Mutation plays a critical role in GA. As discussed earlier, crossover leads the population to converge by making the chromosomes in the population alike. Mutation reintroduces

genetic diversity back into the population and assists the search escape from local optima.

Being a population-based approach, GA are well suited to solve multi-objective optimization problems. A generic single-objective GA can be modified to find a set of multiple non-dominated solutions in a single run. The ability of GA to simultaneously search different regions of a solution space makes it possible to find a diverse set of solutions for difficult problems with non-convex, discontinuous, and multi-modal solutions spaces. The crossover operator of GA may exploit structures of good solutions with respect to different objectives to create new nondominated solutions in unexplored parts of the Pareto front. In addition, most multi-objective GA do not require the user to prioritize, scale, or weigh objectives. Therefore, GA have been the most popular heuristic approach to multi-objective design and optimization problems. Jones etal. [15] Reported that 90% of the approaches to multiobjective optimization aimed to approximate the true Pareto front for the underlying problem. A majority of hese used a meta-heuristic technique, and 70% of all metaheuristics approaches were based on evolutionary approaches.

Several survey papers [16], [17], [18] and [19] have been published on evolutionary multi-objective optimization. Coello lists more than 2000 references in his website [20]. Generally, multi-objective GA differ based on their fitness assignment procedure, elitisim, or diversification approaches. In this paper, multi-objective genetic algorithm to select the optimal number of features to use.

IV. CORRELATION BASED FEATURE SELECTION (CFS)

Like the majority of feature selection programs, CFS uses a search algorithm along with a function to evaluate the merit of feature subsets. The heuristic by which CFS measures the "goodness" of feature subsets takes into account the usefulness of individual features for predicting the class label along with the level of intercorrelation among them. The hypothesis on which the heuristic is based can be stated: " Good feature subsets contain features highly correlated with the class, yet uncorrelated with each other."

In test theory [21], the same principle is used to design a composite test (the sum or average of individual tests) for predicting an external variable of interest. In this situation, the features" are individual tests which measure traits related to the variable of interest (class).

A feature subset S containing k features, $\overline{r_{cf}}$ the average feature – class correlation, and $\overline{r_{ff}}$ the average feature – feature intercorrelation, The equations are defined.

$$\overline{r_{ff}}=\frac{\sum_{i=,}^{n-1}\mathrm{Co}(f_i,f_{i+1})}{n} \quad (19)$$

$$\overline{r_{cf}}=\frac{\sum_{i=1}^{n}\mathrm{Co}(c,f_i)}{n} \quad (20)$$

Where Co is the correlation matrix. In this paper, we compare the number of selected features FSFSM – MOGA method with CFS method using equations (19) and (20).

V. THE PROPOSED FEATURE SELECTION METHOD BASED ON FUZZY SIMILARITY MEASURES BY MULTI OBJECTIVE GENETIC ALGORITHMS

This method is the same as CFS method, With the difference that instead of the correlation matrix use fuzzy

similarity measure matrix and to select the optimal the number of features, we use multi-objective genetic algorithm.

In this case, equations (19) and (20) the following equations are converted.

$$\overline{S_{ff}} = \frac{\sum_{i=0}^{n-1} Sm(f_i, f_{i+1})}{n} \quad (21)$$

$$\overline{S_{cf}} = \frac{\sum_{i=1}^{n} Sm(c, f_i)}{n} \quad (22)$$

Where Sm is the fuzzy similarity measure matrix, $\overline{S_{cf}}$ the average feature – class fuzzy similarity, and $\overline{S_{ff}}$ the average feature – feature fuzzy similarity.

Use MOGA algorithm of In the proposed method is explained briefly as follows:

1. Fuzzy similarity matrix calculated using fuzzy similarity measures.
2. Calculated $(\overline{S_{ff}})$ and $(\overline{S_{cf}})$ using fuzzy similarity matrix.
3. Selection the number of optimal features using multi-objective genetic algorithm (use fitness functions relating to equations (21) and (22)).

$$f(1) = \overline{S_{ff}} = \frac{\sum_{i=0}^{n-1} Sm(f_i, f_{i+1})}{n} \quad (23)$$

$$f(2) = \frac{1}{\overline{S_{cf}}} = \frac{n}{\sum_{i=1}^{n} Sm(c, f_i)} \quad (24)$$

Finally, we compare the results of proposed method with the results of the CFS method fitness functions. CFS method, Fitness functions:

$$f(1) = \overline{r_{ff}} = \frac{\sum_{i=0}^{n-1} Co(f_i, f_{i+1})}{n} \quad (25)$$

$$f(2) = \frac{1}{\overline{r_{cf}}} = \frac{n}{\sum_{i=1}^{n} Co(c, f_i)} \quad (26)$$

VI. EXPERIMENTS

Six data sets are used in the experiments to test the proposed algorithm. All of them have been taken from the UCI machine learning database. [22]. the properties of the data sets are summarized in Table (3) (they differ greatly in the sample size, feature number).

Table (3): Description of the used data sets.

No.	Data sets	Features	Samples
D_1	Semeion Handwritten Digit	266	1593
D_2	Dbworld	4702	64
D_3	Dbworld_bodies_stemmed	3721	64
D_4	Lung – Cancer	52	26
D_5	Madelon	500	1800
D_6	CNAE –9	857	1080

Compare the results proposed method with CFS method come in Table (4).

Table (4): The number of Selective features the proposed method in comparison with the conventional method for ∩ - FSM using correlation matrix.

Data sets	Total number of features	The number of selective feature two method	
		CFS method	FSFSM – MOGA method
		Using the correlation matrix	∩ - FSM
D_1	266	12	10
D_2	4702	260	138
D_3	3721	207	106
D_4	52	2	2
D_5	500	32	26
D_6	857	36	24

In the feature selection algorithms of machine learning, the average fuzzy measure feature – feature less and average fuzzy measure class – feature more is better. In table (4), we see that the FSFSM – MOGA method is better than the conventional method. Especially when we use of the ∩ - fuzzy similarity measures. This result in Figure (2) is clearly specified.
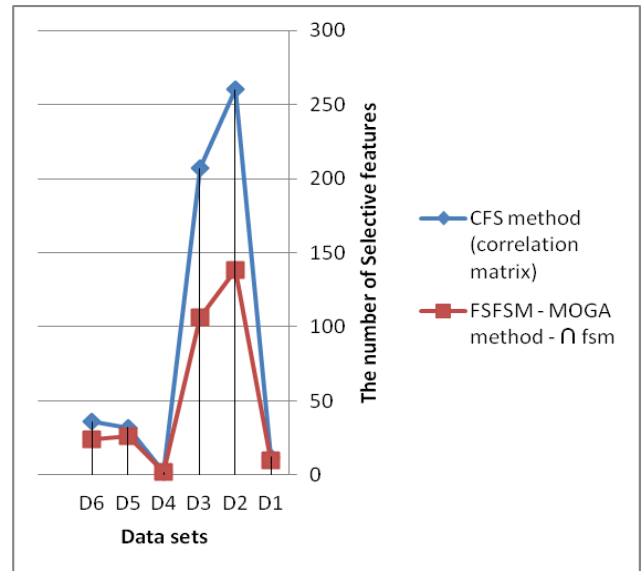

Figure 2. The number of Selective features the proposed method in comparison with CFS method.

VII. CONCLUSION

In this paper, we have presented a new method for feature selection based on ∩ - fuzzy similarity measure using multi – objective genetic algorithm (FSFSM – MOGA). CFS method using ∩ - fuzzy similarity measure and MOGA were improved, as the performance of this method is shown in Table (4) and figure (2). Experiments show that feature selection method using fuzzy similarity measures together with similarity classifier is giving good results. The proposed method can be modified with other fuzzy similarity measures and improve it.

REFERENCES

[1] M. Dash and H. Liu, "Feature Selection for Classification," (1997): Intelligent Data Analysis, 1(1 – 4): 131 – 156.
[2] Kaufman L, Rousseeuw PJ. , (1990): Analysis New York: Wiley.
[3] Zadeh LA. Fuzzy sets. (1965): Information Control 8:338–356.
[4] Zwick R, Carlstein E, Budescu DV. Measures of similarity among fuzzy concepts: A comparative analysis. (1987): Int J Approximate Reason 1:221–242.
[5] Pappis CP, Karacapilidis NI. A comparative assessment of measures of similarity of fuzzy values. (1993): Fuzzy Sets and Systems 56:171–174.
[6] Wang WJ. New similarity measures on fuzzy sets and on elements. (1997): Fuzzy Sets and Systems 85:305–309.
[7] Liu X. Entropy, distance measure and similarity measure of fuzzy sets and their relations. (1992): Fuzzy Sets and Systems 52:305–318.
[8] Fan J, Xie W. Some notes on similarity measure and proximity measure. Fuzzy Sets Syst 1999; 101:403–412.
[9] Turksen IB, Zhong Z. An approximate analogical reasoning approach based on similarity measures. (1988): IEEE Trans Syst Man Cybern 18:1049–1056.

[10] Buckley JJ, Hayashi Y. Fuzzy input–output controllers are universal approximates. (1993): Fuzzy Sets and Systems 58:273–278.

[11] Kalle Saastamoinen (2008): many valued algebraic structure as measures of comparison. PhD thesis, Lappenranta University of Technology.

[12] Gottwald Siegfried (1993): Fuzzy sets and fuzzy logic. Arti_cial intelligence, ISNB 3-528-05311-9.

[13] Bustince, H. Indicator of inclusion grade for interval valued fuzzy sets: Application to approximate reasoning based on interval-valued fuzzy sets, (2000): Int. J. Approx. Reasoning, 23: 137-209.

[14] De Baets, Bernard, and Hans De Meyer., "The Frank T – norm Family in Fuzzy Similarity Measurement.", (2001): In EUSFLAT Conference, 2nd, Proceedings,249–252.

[15] Jones DF, Mirrazavi SK, Tamiz M. Multiobjective meta-heuristics: an overview of the current state-of-the-art. (2002):Eur J Oper Res;137(1):1–9.

[16] Zitzler E, Deb K, Thiele L. Comparison of multiobjective evolutionary algorithms: empirical results. (2000): Evol Comput; 8(2):173–95.

[17] Zitzler E, Thiele L. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. (1999): IEEE Trans Evol Comput; 3(4):257–71.

[18] Coello CAC. An updated survey of GA-based multiobjective optimization techniques. (2000): ACM Comput Surv; 32(2):109–43.

[19] Jensen MT. Reducing the run-time complexity of multiobjective EAs: The NSGA-II and other algorithms. (2003): IEEE Trans Evol Comput; 7(5):503–15.

[20] Xiujuan L, Zhongke S. Overview of multi-objective optimization methods. (2004): J Syst Eng Electron;15(2):142–6.

[21] Huawen Liu, JiguiSun, LeiLiu, HuijieZhang, Feature selection with dynamic mutual information. (2009): Pattern Recognition 42: 1330 – 1339.

[22] Newman, D. J., Hettich, S., Blake, C. L., & Merz, C. J, (2007): UCI Repository of machine learning databases. Irvine, CA: University of California, Department of Information and Computer Science.