# Processing measure uncertainty into fuzzy classifier

**Thomas Monrousseau**[1]**, Louise Travé-Massuyès**[1] and **Marie-Véronique Le Lann**[1,2]

[1]CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France

[2]Univ de Toulouse, INSA, LAAS, F-31400 Toulouse, France

e-mails: thomas.monrousseau@laas.fr, louise@laas.fr ,mvlelann@laas.fr

## Abstract

Machine learning such as data based classification is a diagnosis solution useful to monitor complex systems when designing a model is a long and expensive process. When used for process monitoring the processed data are available thanks to sensors. But in many situations it is hard to get an exact measure from these sensors. Indeed measure is done with a lot of noise that can be caused by the environment, a bad use of the sensor or even the conversion from analogic to numerical measure. In this paper we propose a framework based on a fuzzy logic classifier to model the uncertainty on the data by the use of crisp (non fuzzy) or fuzzy intervals. Our objective is to increase the number of good classification results in the presence of noisy data. The classifier is named LAMDA (Learning Algorithm for Multivariate Data Analysis) and can perform machine learning and clustering on different kind of data like numerical values, symbols or interval values.

## 1 Introduction

Data classification is the process of dividing pattern space using hard, fuzzy or probabilistic partitions into a number of regions [1]. Classification algorithms are more and more used nowadays in a world where it is not always simple to get a model of complex process. On the opposite it is easier to get data on systems by monitoring and store it. Different types of classifiers can be used depending on the situation. The principal ones described in the literature are artificial neural networks, k-nearest neighbors, support vector machine, decision trees, fuzzy classifiers and statistical methods.

Most of the time, data are issued from sensor measurements and are corrupted by noise. This noise can have different origins, for example environment disturbances, bad use of the sensor, hysteresis effect or numerical conversion and representation of the data. Many domains of application have to deal with noise problems like medical diagnosis [2], biologic identifications [3] or image recognition [4]. Uncertainty can be understood in two ways: the first is the uncertainty directly present in the data like noise and the second can be assimilated as the reliability of a feature inside a class. In this paper we consider only the first case. To avoid noise problems in classification some solutions have been provided previously, for example the transformation of

data [5] [6] [7], the use of fuzzy logic type-1 or type-2 [3] or statistical models.

Fuzzy logic is a multi-valued logic framework introduced by Zadeh [8] that is known to be more efficient for representating uncertainty and impreciseness than binary logic. In previous work, a fuzzy classifier named Learning Algorithm for Multivariate Data Analysis (LAMDA) has been proposed by Aguilar [9]. This classifier can originally process simultaneously two different types of data: quantitative data and qualitative data. A real number contains an infinite amount of precision whereas human knowledge is finite and discrete, thus LAMDA is interesting because there is no solution proposed in the literature to process in a uniform way heterogeneous data and to handle in a same problem quantitative data and qualitative data is often a complex subject. A new type of data, the interval, has been introduced by Hedjazi [10] to model uncertainties by means of crisp intervals. In this paper we propose an extention to fuzzy intervals in order to improve its application to process noisy data measurements but with the capacity to handle others features types like "clean" data or qualitative features. Moreover the algorithm should stay low cost in term of memory and computation time to enable the method to be embedded on small systems.

In the first part of the paper the LAMDA algorithm is shortly presented then in a second time a method to use the algorithm to classify noisy data is introduced. This method is in two parts: the first presents a general solution to model uncertainty on data with crisp intervals based on confidence intervals and the second shows an improvement to model Gaussian noise with fuzzy intervals. In both cases examples of application are introduced to show the improvement of the method compared to the use of the data without transformation.

## 2 LAMDA algorithm (Learning Algorithm for Multivariate Data Analysis)

This section presents the principle of the LAMDA algorithm.

### 2.1 General principle

LAMDA is a classification algorithm based on fuzzy logic created on an original idea of Aguilar [9] and can achieve machine learning and clustering on large data sets.

The algorithm takes as input a sample $x$ made up of $N$ features. The first step is to compute for each feature of $x$, an
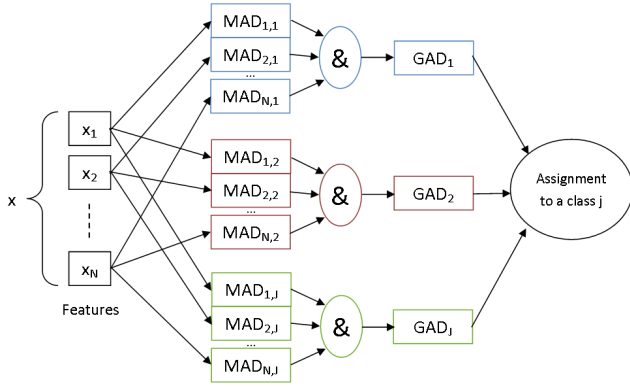
Figure 1: Summarized scheme of the LAMDA algorithm

adequacy degree to each class $C_j$, $j = 1..J$ where $J$ is the total number of class. This is obtained by the use of a fuzzy adequacy function. So $J$ vectors of $N$ adequacy degrees are computed, these vectors are called Marginal Adequacy Degree vectors (MAD). At this point, all the features are in a common space. Then the second step is to take all the MADs and aggregate them into one global adequacy degree (GAD) by means of a fuzzy aggregation function. Thus the $J$ MAD vectors (composed of $N$ MADs) become $J$ scalar GADs, the higher the GAD, the better the adequacy to the class. The simplest way to assign the sample $x$ to a class is to keep as result the class with the biggest GAD.

All the process is summarized in Fig. 1.

## 2.2 Fuzzy membership computation

During the learning step, the algorithm creates prototype data for each class and for each feature. These data are called classe descriptors or prototypes; they can be for example means or variances. We define as $C_{j,n}$ the class prototype of the *n-th* feature for the class $j$.

As previously mentioned the first step of the algorithm is a comparison between the sample vector $x$ and all the $C_{j,n}$. This operation is performed with membership functions and gives as result a membership adequacy degree. Thus $MAD_{j,n}$ is the MAD for the *j-th* class and the *n-th* feature. As the framework is based on fuzzy logic, all memberships are numbers in the [0,1] interval. The general membership function is:

$$MAD_{j,n} = f(C_{j,n}, x_n) \qquad (1)$$

The class prototype $C_{j,n}$ depends on two things: the type of data and the function used. Some functions may require only one data into $C_{j,n}$ whereas others need a list of parameters.

In the following section, some examples of membership functions are presented.

- Quantitative data:

  Many functions are available for this kind of data. For example the Gaussian:

$$f(x_n) = e^{-\dfrac{(x_n - \rho_{j,n})^2}{2\sigma_{j,n}^2}} \qquad (2)$$

  or the binomial function:

$$f(x_n) = \rho_{j,n}^{x_n} \cdot (1 - \rho_{j,n})^{1-x_n} \qquad (3)$$

Where $x_n$ is the *n-th* feature of the sample $x$, $\rho_{j,n}$ is the mean of the *n-th* feature for the class $j$ and $\sigma_{j,n}$ is the standard deviation of the *n-th* feature for the class $j$.

- Qualitative data:

  Qualitative can take values in a set of modalities. The membership function of qualitative data returns the frequency of modality taken by the feature into the class during the learning phase. We introduce a qualitative variable with $K$ modality $\{Q^1, ..., Q^K\}$ and the frequency $\Phi_j^k$ of the modality $Q^k$ for the class $j$. The membership is described by:

$$f(x_n) = (\Phi_{j,n}^1)^{q_1} * ... * (\Phi_{j,n}^K)^{q_K} \qquad (4)$$

$$\text{with } \begin{cases} q^k = 0 \text{ if } x_n \neq Q_k \\ q^k = 1 \text{ if } x_n = Q_k \end{cases}$$

- Intervals:

  The membership function for interval data is a function which tests the similarity between two fuzzy intervals. In this case similarity is defined by two components: the distance between the intervals and the surface that these intervals have in common. Indeed the class prototype for crisp interval data is a mean interval. The similarity function is:

$$S(A,B) = \frac{1}{2}\left(\frac{\int_V \mu_{A \cap B}(\xi)\mathrm{d}\xi}{\int_V \mu_{A \cup B}(\xi)\mathrm{d}\xi} + 1 - \frac{\partial[A,B]}{\varpi[V]}\right) \quad (5)$$

where $\mu_X(x)$ is the value of $x$ in the fuzzy set $X$, $\partial[A,B]$ is the distance between intervals A $= [a^-, a^+]$ and B $= [b^-, b^+]$ and $\varpi[X]$ is the size of a fuzzy set into a V universe. This is described by:

$$\varpi[X] = \int_V \mu_X(\xi)\mathrm{d}\xi \qquad (6)$$

In the case of crisp intervals and in a universe between 0 and 1:

$$S(A,B) = \frac{1}{2}\left(\frac{\varpi[A \cap B]}{\varpi[A \cup B]} + 1 - \partial[A,B]\right) \qquad (7)$$

where $\varpi[X]$ in this case can be replaced by the length of the interval:

$$\varpi[X] = upperbound(X) \text{-} lowerbound(X) \qquad (8)$$

and distance $\partial[A,B]$ is defined as:

$$\partial[A,B] = max[0, max(a^-, b^-) - min(a^+, b^+)] \quad (9)$$

In the case where an interval feature is used the prototype for a class $j$ is given by $[\rho_j^{n-}, \rho_j^{n+}]$ where $\rho_j^{n-}$, respectively $\rho_j^{n+}$ represents the mean value of lower bounds (respectively upper bounds) of all the elements belonging to class j for this feature.

Once the MAD are computed whatever the feature type, it is possible to perform any type of processing as described on Fig. 2.
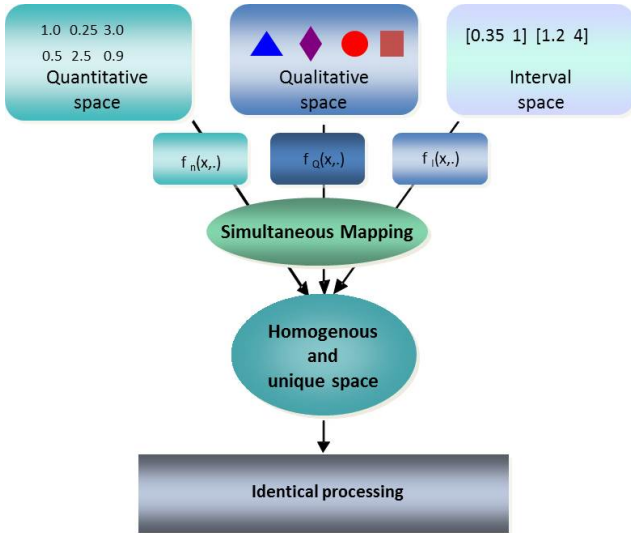
Figure 2: Projection principle for heterogeneous feature types

## 2.3 Marginal adequacy degree merging

Once all the features are grouped into the membership space the next step of the algorithm is to transform the MAD vectors into a set of single value which depicts the global membership of the sample to a class. These values were introduced in section 2.1 and are called GAD. To perform this transformation a fuzzy aggregation function $\Psi$ is used.

The aggregation function is the following:

$$\Psi(MAD) = \alpha.\gamma(MAD) + (1 - \alpha).\beta(MAD) \quad (10)$$

where $\gamma$ is a fuzzy T-norm and $\beta$ is a fuzzy T-conorm. $\alpha$ parameter is called exigency indicator. It enables to give more or less significance to the union operation and the intersection operation. Two fuzzy T-norm and T-conorm are currently implemented in the algorithm, the min-max and the probabilistic. For example if min-max is used, (10) becomes:

$$\Psi(MAD) = \alpha.min(MAD) + (1 - \alpha).max(MAD) \quad (11)$$

When all GAD are computed they give the membership of the data $x$ to each class. The final result depends on the application but the simplest way to give a result is to class the sample in the class which has the highest GAD. A limit membership can also be fixed: if no GAD is higher than the limit, the sample is defined as unclassifiable.

## 3 Uncertainty modeled with crisp intervals

### 3.1 Method presentation

Every data measurement is performed with noise. In some cases noise has enough bad effect to increase the error of classification. Thus the point is to model the imprecision of the data to decrease the number of bad classifications.

A technique used in several fields of application is the use of intervals to symbolize data uncertainty [11] [12]. So we are suggesting a framework where numerical data are transformed into intervals to model imprecision.

In a situation where the probability law followed by the noise on a variable is unknown, it may be possible to obtain a confidence interval. It is an interval in which the real value of the measure is present with a certain amount

of confidence (for example a confidence interval of 95% is an interval in which the exact value of the measure can be found with a probability of 95%). Introducing $\hat{x}$ the measured value and $l$ the length of a centered on zero confidence interval based on the measurement error, the interval used by the algorithm is calculated: $X = [\hat{x} - \frac{l}{2}; \hat{x} + \frac{l}{2}]$.

The main aim of the transformation is to improve the classification on the transition zones where data is really sensitive to noise and a small change can modify the output of the classifier. The use of intervals to model uncertainty is effective only if the "clean" data is relevant for the classification problem. If it is not the case a better solution is to remove the irrelevant feature. It will in most cases provide better output results. This expresses the fact that if the "clean" data is difficult to classify it is not improved by using confidence intervals.

### 3.2 Experiments

A set of data has been created for an application test which can be interpreted as sensors time evolution of a continuous process. This set of data is composed by three quantitative (numerical) features of 101 samples that are shown on the Fig. 3. Three classes are specified and used as targets for the classifier. These classes are chosen arbitrarily to represent different behaviors of a system that could be healthy or failure modes. Nevertheless the classes are built to make all the data relevant for the system monitoring which means the three features do not have a global negative impact on the classification results.

The three features $x, y$ and $z$ are defined by the following time functions:

- $x = e^{\frac{-t}{2}}$
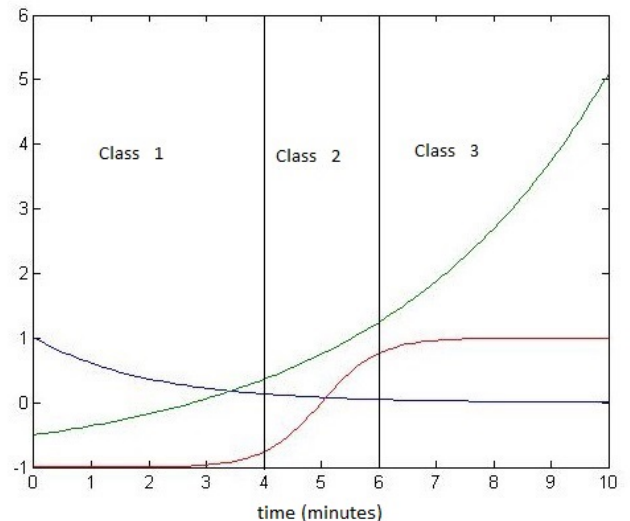- $y = \frac{1}{2} \cdot e^{\frac{t}{4}} - 1$
- $z = tanh(t - 5)$



Figure 3: Data used to test the intervals method

This example is used to measure the improvement in the classification results in the case of all data are noisy. Artificial noise is added by the following: $x$ is the ideal variable without noise and $\hat{x}$ the noisy variable, $\hat{x} = x + Y$ with
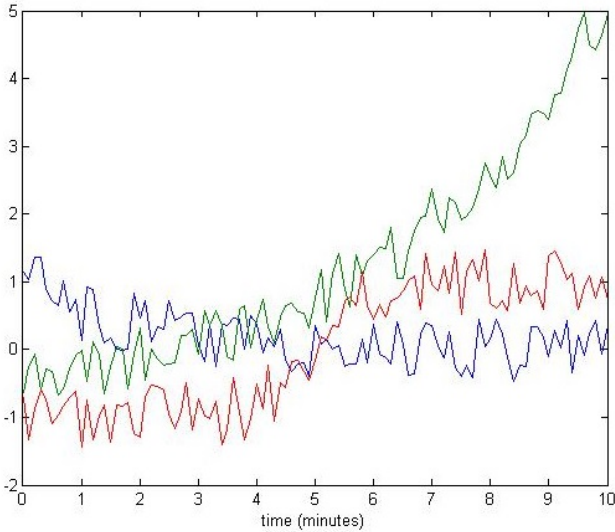
Figure 4: An example of data corrupted with a noise in the interval [-0.5 ; 0.5]

$Y$ a random variable following a uniform distribution on an interval $I$.

The experiment has been performed with these conditions: $\alpha$ parameter of (10) is set at 0.8 with the [min,max] functions to compute the fuzzy aggregation and the membership function used for quantitative data is the binomial.[min, max] aggregation is chosen because experiments on the algorithm showed that this kind of aggregation provides better results on noisy data that the probabilistic one. A first classification without any noise gives a result of 91% of good classification. Then the experiment is repeated a great many times to avoid statistical mistakes. In this case, the experiment has been run fifty thousand times, $\hat{x}$ is re-computed at each new run. Results are given on table 1.

| Interval for random data | [-0.3 ; 0.3] | [-0.5 ; 0.5] | [-2 ; 2] |
|---|---|---|---|
| Mean success percentage with binomial function | 89.9% | 84.7% | 79.6% |
| Mean success percentage with interval function | 91.9% | 89.8% | 70.3% |

Table 1: Table of results for the crisp intervals method

As it can be seen, this method provides an improvement on the results in the two first cases where noise deteriorates the classification with the quantitative method but when the data is still globally consistent. In these cases, the intervals method gives better results than binomial method 82% of the time. But when noise amplitude is much higher than the data like in the $[-2; +2]$ error interval, the interval method does worse in general than the binomial function.
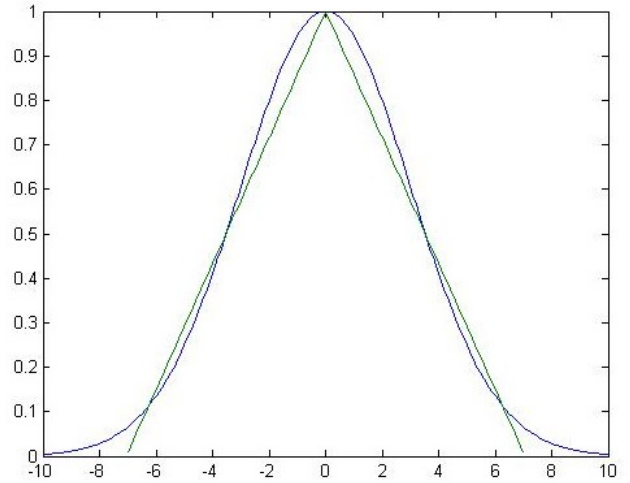


Figure 5: Example of approximation of a Gaussian fuzzy interval by a triangular fuzzy interval

## 4 Modeling Gaussian noise with fuzzy intervals

### 4.1 Fuzzy interval method presentation

Most of the time, noise on physical measure follows a Gaussian distribution centered on the real value. Thus it is interesting to model this specific kind of uncertainty. Nevertheless, it is difficult to handle fuzzy intervals with an exact Gaussian shape. That is why we suggest approximating the Gaussian with a triangular fuzzy interval. This interval is described with a lower boundary $x^-$ and an upper boundary $x^+$: $X = [x^-; x^+]$ which leads to a similar description as crisp intervals. So:

$$\mu_X(x^-) = 0 \text{ and } \mu_X(x^+) = 0 \text{ and } \mu_X(\tfrac{x^+ + x^-}{2}) = 1$$

with $\mu_X(x)$ the fuzzy value of $x$ into the fuzzy set $X$. As a Gaussian of $\rho$ mean is centered on the true measure value the maximum fuzzy value of the triangle $\frac{x^+ + x^-}{2}$ is equal to $\rho$. To compute $x^-$ and $x^+$ we propose to use the full width at half maximum (FWHM) that can be calculated this way:

$$FWHM = 2\sqrt{2ln(2)} \cdot \sigma \qquad (12)$$

with $\sigma$ that is the standard deviation of the measure. Thus for a Gaussian function that has a mean value $\rho$ and a standard deviation $\sigma$ the approximated interval $X$ is defined by $X = [\rho - 2\sqrt{2ln(2)} \cdot \sigma; \rho + 2\sqrt{2ln(2)} \cdot \sigma]$. An example of this approximation is given on Fig. 5.

Until now all the implementations of the LAMDA algorithm were using only crisp intervals despite the fact that the general method was introduced. The class prototype is now a triangle interval computed with the means of upper and lower boundaries of the data used to train the algorithm. Thus the membership function is still a similarity measure between two fuzzy intervals like in (5) but it is necessary to redefine the distance function between the intervals. A solution has been proposed to measure a distance with the center of gravity of triangular fuzzy intervals [13]. In the present situation:

$$\partial[A, B] = |\frac{a_+ + a_-}{2} - \frac{b_+ + b_-}{2}| \qquad (13)$$

with $A = [a_-; a_+]$ and $B = [b_-; b_+]$, $A$ and $B$ being triangular fuzzy intervals like described in this section.

The intersection $A \cap B$ needed in (5) is calculated with an analytical solution based on geometry and trigonometry. It avoids numerical integration that could be less precise and longer to compute.

## 4.2 Experiments

As we did previously with the crisp method, a test is performed with a Gaussian noise on the same data set (Fig. 3). The test is done in the same conditions as in the previous section. The difference is on the construction of the noisy data $\hat{x} = x + Y$. $Y$ is now a random variable that follows a normal distribution of standard deviation $\sigma$ and centered on 0. Results of the simulation are given on the table 2.

| $\sigma$ | 0.2 | 0.5 | 0.7 | 1 |
|---|---|---|---|---|
| Mean success percentage with binomial function | 83.2% | 79.8% | 79.8% | 79.6% |
| Mean success percentage with crisp interval function | 86.8% | 82.5% | 77.2% | 71.3% |
| Mean success percentage with fuzzy interval function | 93.1% | 84.5% | 79.3% | 74.8% |

Table 2: Table of results for the fuzzy intervals method

Similarly to the previous test, the interval method increases the rate of good classifications until the standard deviation $\sigma$ becomes too high and the binomial function provides better results. This point is reached here for $\sigma = 0.7$ which corresponds to a signal to noise ratio (SNR) of 6 dB for the signal with the smallest amplitude. Also it is important to notify that in all cases the fuzzy interval provides better results than the crisp interval method.

## 4.3 Experiments on iris dataset

As a second example we use the classical iris dataset[14]. This dataset contains four features: sepal length in cm, sepal width in cm, petals length in cm and petal width in cm. All these features are measured for three types of flower: iris Setosa, iris Versicolour and iris Virginica which constitute three classes. It is easy to classify without any error the iris dataset by using only the petals information that are in general most relevant that the sepals ones. Thus only the sepal sizes are kept in this test to simulate the noise. The figure 6 shows the repartition of the data in the 2D space of the sepal features.

We assume that the data follow a normal distribution centered on a mean $\mu_{j,n}$ and with a standard-deviation $\sigma_{j,n}$. This hypothesis can be verified by using a statistical test. The Kolmogorov-Smirnov test has been used for each class with a 5% significance level, it shows that the hypothesis is true for the iris Setosa and the iris Versicolour but not for the iris Virginica. Nevertheless all the data are processed as if they follow a normal distribution.
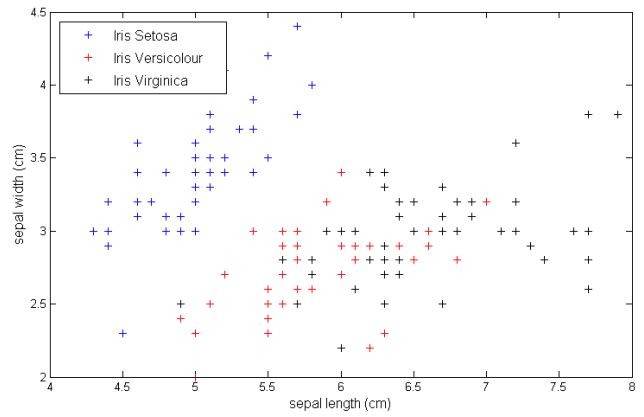


Figure 6: Representation of iris data by class

The classifications are performed using the cross-validation method. The percentages of well classified data for the two methods are:

- using binomial function (scalar): 81.3%
- using fuzzy triangular intervals: 94.0%

Once again the classification rate is increased by the use of the fuzzy interval method instead of the binomial one.

## 5 Conclusion

We presented in this article two methods to model uncertainty for classification applications. An example showed that these methods can improve classification results even when the signal to noise ratio is high. The second method based on fuzzy intervals demonstrated that try to model more precisely the probability law of the noise can provide better results than use confidence intervals modelled by crisp intervals. However this process to model uncertainty reveals limits when the SNR reaches a low level. A future important work is to limit the classification error of the interval method at the level of the numerical method.

These methods will now be tested on data out coming from a real industrial process.

Another way to manage uncertainty on classifiers like LAMDA could be to use type-2 fuzzy functions [15]. This is an expansion of classical fuzzy logic where the membership functions give in output a fuzzy interval which can be used to model variance of the data.

To provide a better solution to manage uncertainty in the LAMDA classifier it can be useful to extend the problem to the qualitative features. It is often difficult to determine if a qualitative element is close to another, for example the color "orange" is closer to "red" than "blue". But on small training dataset consider this kind of information can improve final classification results. This could be done by using similarity matrix which are already used in some artificial intelligence problems.

LAMDA algorithm can work with a feature selection algorithm named MEMBAS (Membership Margin Based Feature Selection) [16]. This algorithm uses LAMDA classes definitions and its membership functions to provide an analytical solution for the feature selection. A future work will be to measure the impact of the interval use on MEMBAS algorithm to perform selection on noisy data.

# References

[1] J. C. Bezdek. A review of probabilistic, fuzzy, and neural models for pattern recognition. *Journal of Intelligent and Fuzzy Systems*, Vol. 1, No. 1:pp 1–25, 1993.

[2] E. Alba, J. Garcia-Nieto, L. Jourdan, and E. Talbi. Gene selection in cancer classification using pso/svm and ga/svm hybrid algorithms. In *Evolutionary Computation, CEC 2007. IEEE Congress on*, pages 284–290, Sept. 2007.

[3] Scott Ferson, H. Resit Akqakaya, and Amy Dunham. Using fuzzy intervals to represent measurement error and scientific uncertainty in endangered species classification. In *Fuzzy Information Processing Society, 1999. NAFIPS. 18th International Conference of the North American on*, pages pp 690–694, Jul 1999.

[4] Zhang Weiyu, S.X. Yu, and Shang-Hua Teng. Power svm: Generalization with exemplar classification uncertainty. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages pp 2144–2151, June 2012.

[5] Arafat Samer, Dohrmann Mary, and Skubic Marjorie. Classification of coronary artery disease stress ecgs using uncertainty modeling. In *Computational Intelligence Methods and Applications, 2005 ICSC Congress*, 2005.

[6] Kynan E. Graves and Romesh Nagarajah. Uncertainty estimation using fuzzy measures for multiclass classification. *Neural Networks, IEEE Transactions on*, Vol. 18:pp. 128–140, 2007.

[7] Prabha Verma and R.D.S. Yadava. Fuzzy c-means clustering based uncertainty measure for sample weighting boosts pattern classification efficiency. In *Computational Intelligence and Signal Processing (CISP), 2012 2nd National Conference on*, pages 31–35, 2012.

[8] L.A. Zadeh. Fuzzy sets. *Information and Control*, vol. 8:pp. 338–353, June 1965.

[9] Carrete N.P. and Aguilar-Martin J. Controlling selectivity in nonstandard pattern recognition algorithms. In *IEEE Transactions on Systems, Man and Cybernetics*, volume 21, pages 71–82. IEEE, Jan/Feb 1991.

[10] Hedjazi L., Aguilar-Martin J., Le Lann M.V., and Kempowsky T. Towards a unfined principle for reasoning about heterogeneous data: a fuzzy logic framework. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 20, No. 2:pp. 281–302, 2012.

[11] B. Kuipers. *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge*. The MIT Press,Cambridge, Massachusetts, london edition, 1994.

[12] Lynne Billard. Some analyses of interval data. *Journal of Computing and Information Technology*, CIT 16:pp 225–233, 2008.

[13] Hsieh C. H. and Chen S. H. Similarity of generalized fuzzy numbers with graded mean integration representation. In *Proceedings of the Eighth International Fuzzy Systems Association World Congress*, volume vol. 2, pages pp. 551–555, Taipei, Taiwan, Republic of China, 1999.

[14] Fisher R.A. {UCI} machine learning repository, 1936. http://archive.ics.uci.edu/ml.

[15] J.M. Mendel, R.I. John, and F. Liu. Interval type-2 fuzzy logic systems made simple. *Fuzzy Systems, IEEE Trans. on*, Vol. 14, No. 6:pp 808–821, Dec. 2006.

[16] L.Hedjazi, J.Aguilar-Martin, and M.V. Le Lann. Similarity-margin based feature selection for symbolic interval data. *Pattern Recognition Letters*, Vol.32, No4:pp. 578–585, March 2012.