Fuzzy Output Error

Tom Gedeon, Leana Copeland, Sumudu Mendis

Research School of Computer Science
Australian National University
E-mail: {tom.gedeon | leana.copeland | sumudu.mendis}@anu.edu.au

Abstract  Many training algorithms use mean square error (MSE) for training when we have numeric data, with MSE also used to indicate the quality of the results. In many real world applications we are also interested in the number of samples which are “close enough” to the correct value and would also use the number of samples which could be mapped to correct classifications in reporting our results. We extend our previous work to use fully fuzzy output values, and demonstrate the benefits of our approach which allows choice in the shape of fuzzy output error membership functions which can be used to specialize the approach to particular domains or adapt to particular kinds of data sets. We show how we can extend our approach using fuzzy true positive, fuzzy true negative, fuzzy false positive, fuzzy false negative identifications as well as regression line weighted variants of these.

Previous work  Fuzzy Classification Error  In our previous work (Mendis and Gedeon, 2008) we formulated the Sum of Fuzzy Classification Error (SYCLE) in the following way. We call it Fuzzy as it considers transition between good and bad classifications using several categories of error. First, we specify that both desired output and predicted output of an experiment are in the range [0, 1]. Next, we define a set of rules for the classification and these rules are visualised in the following figure.

According to Figure 1, there are 3 categories of classifications that can occur, they are Good, Bad, and Very Bad (V. Bad). Now we assume the pair of predicted and desired values, of the ith input, respectively taken as X and Y coordinates of the point \( P_i \) on the 2 dimensional fuzzy classification error rule space, Figure 1. The fuzzy classification error of an arbitrary point \( P_i \) can be written as,

\[
FYCLE(P_i) = \begin{cases} 
0 & \text{if } P_i \text{ Good} \\
0.5 & \text{if } P_i \text{ Bad} \\
1 & \text{if } P_i \text{ VeryBad} 
\end{cases}
\]

Let us consider the 4 straight lines, B1, B2, G1, and G2, in Figure 1. In this experiment, they are equivalent to,

\[ B1 \equiv y-x-0.5 \quad G1 \equiv y-x-0.2 \quad G2 \equiv y-x+0.2 \quad B2 \equiv y-x+0.5 \]

Now, The fuzzy classification error of an arbitrary point \( P_i \) can be calculated as,


\[ \text{FYCLE}(P) = \begin{cases} 
0 & \text{if } G_1(P) \leq 0 \text{ AND } G_2(P) \geq 0 \\
0.5 & \text{if } (B_1(P) \leq 0 \text{ AND } G_1(P) > 0) \text{ OR } (G_2(P) < 0 \text{ AND } B_1(P) > 0) \\
1 & \text{if } B_1(P) > 0 \text{ AND } B_2(P) < 0 
\end{cases} \]

Next, the Sum of Fuzzy Classification Error (SYCLE) for a set of data with \( m \) records can be calculated as:

\[ \text{SYCLE} = \frac{1}{m} \sum_{i=1}^{m} \text{FYCLE}(P_i) \]

where \( m \)

The problem with our previous approach is that while it is multi-valued beyond 0 and 1, it is not fuzzy, as we use only the three discrete values of 0, 0.5 and 1. In the next section we introduce our extensible approach to this problem.

**Fuzzy output error**

In principle, our approach is straightforward and consists of superimposing a fuzzy membership function onto the contents of Figure 1. This can be seen in Figure 2.

The *Good* region retains goodness value of 1, and *Very Bad* regions 0, but the *Bad* region is not limited to just 0.5 but has a value from 0 to 1 depending on the data location between the lines B2 to G2 or B1 to G1.

We can extract the shape of a fuzzy membership function using the projections onto the axes for new axis values as follows: the *True* values map to negative, while actual *Output* values map to positive, see Figure 3. This is an arbitrary convention to match the shape of the trapezoid used in Figure 2.
Please note the labeling convention on the universe of discourse axis retains absolute values of the original axes. Clearly this is not literally true by Pythagoras’ theorem (Pythagoras, c. 500 BC) since we use diagonal of Figure 2.

Data set

High Salary Selection Problem We select the High Salary Selection problem, as discussed in Gedeon et al (2001), for our experiment. The problem is to find the degree of relevance for having a high salary based on the contacts, age, and work experience of an employee. Figure 4 shows a High Salary Selection Fuzzy Signature, which is obtained using domain expert knowledge, for the high salary selection problem, as used in our previous work (Gedeon et al, 2008).

Fig. 4: High Salary Selection Fuzzy Signature

Note that $\oplus_i$ and $w_i$ in Figure 5 represent the aggregation function and weighted relevance of the node $i$, respectively.

Results

Figure 5 shows the results of normal training using simple gradient descent on mean square error. Most points lie in the Good region, with a substantial minority in the Bad regions which are assigned values of 0.5 in our previous technique. With our new technique, many of these points would, by inspection, be assigned
values greater than 0.5 as being closer to the *Good* region than the respective *Very Bad* regions, hence the numerical error value will be incidentally smaller.

In Figure 5, the MSE is 0.052, the SYCLE is 23, and the FOE₃ is 21.9, the latter using the fuzzy membership curve in Figure 3. The following diagram, Figure 6, shows the relationship of the Data values to Membership values.

Figure 6 also shows the FOE₃ and FOE₈ values we will need later.

We then trained a high salary fuzzy signature on FOE₃ values instead of MSE.

---

**Fig. 5: Test: Fuzzy Classification Error**
– High Salary Fuzzy Signature

**Fig. 6: Expected membership values, MSE trained**
The results on FOE\textsubscript{3} training are encouraging, the MSE has increased but the FOE\textsubscript{3} has decreased. The fuzzy signature is now also no longer trying to learn 0 and 1 values (actually closer to 0.92 in Figure 6), but is learning 0.1 and 0.83 values as being close enough by the FOE\textsubscript{3} central part of the trapezoid.

We tried a number of other fuzzy membership functions with varying results similar in general to Figure 7 above. We reproduce the best result below, using a simplification of the FOE\textsubscript{3} membership function, shown below in Figure 8.

The effect of the membership function in Figure 8 is to assign some degree of correctness to essentially all points, which has the effect of reducing the calculated error using this FOE\textsubscript{8} measure. The value of FOE\textsubscript{8} is 9.2 for the MSE trained case as indicated in Figure 6. The interesting result is from the use of FOE\textsubscript{8} in training, as shown in Figure 9.

The FOE\textsubscript{8} value is reduced by a factor of eight, and the data is clearly

---

**Fig. 7: Expected membership values, FOE\textsubscript{3} trained**

![Graph showing expected membership values, FOE\textsubscript{3} trained](image)

MSE = 0.072
FOE\textsubscript{3} = 14.7

**Fig. 8: Goodness membership function for Fuzzy Output Error, FOE\textsubscript{8}**

![Graph showing goodness membership function, FOE\textsubscript{8}](image)

**Fig. 9: Results of using FOE\textsubscript{8} in training**
much better learnt by inspection of Figure 9. Surprisingly, the MSE is also substantially reduced.

**Extending our approach**

The use of our fuzzy output error approach can be extended to classification problems, Figure 10 shows the 4 classical quadrants for sensitivity and specificity, or terms precision and recall in information retrieval.

The importance of each quadrant depends on the objective in practice. For example, in a medical condition with high mortality, False Positives (Type I errors) are to be preferred over False Negatives (Type II errors). For an annoying condition for which the treatment has side effects, the opposite may be true.
With our Fuzzy Output Error approach, we will prefer to specify some boundary at the edge of each quadrant to indicate the region of uncertainty of the result, given that in practice most of the information used to fill the regions does not in fact come from crisp processes producing only 0s and 1s, but is likely to be a numerical result.

This is shown in Figure 11 with the inner regions corresponding to a goodness of 1 as before.

A similar notion of Fuzzy True Positive etc has been used by Apolloni et al (2006) and Hamlich and Ramdani (2012).

**Combined FOE** The Fuzzy Output Error goodness values shown in Figures 11 and 1 can be combined by intersection, as shown in Figure 12. In Figure 13 we show a similar diagram for really bad prediction results, where the output is strongly a Fuzzy False Positive or Fuzzy False Negative.

We comment that Figures 12 and 13 have been constructed using the core and support of the fuzzy sets represented in Figures 1 and 11 only. If we wish to take into account the 0.5 $\alpha$-cut, or apply a principle of symmetry, two other possible alternatives can be readily identified, as shown in Figure 14.
Our future work will include experimental investigation of extensions of our FOE technique.

**Conclusion**

We have shown that our approach using fuzzy output error can improve the performance of our fuzzy signatures for the high salary selection data set. We have also discussed extension of our idea to the classical binary classification table and illustrated how the integration of quadrant based and regression line based fuzzy output error can be done. Future work will include further investigation of these and the application of our techniques to other data sets, and comparison to other alternative techniques.

**References**


Pythagoras, (c. 500 BC), see also en.wikipedia.org/wiki/Pythagoras_theorem, accessed 12 November 2012.