Comparison of the Use of Different Similarities Based on T-norms in the Classification Tasks

Kalle Saastamoinen
Laboratory of Applied Mathematics
Lappeenranta University of Technology
P.O. Box 20, FIN-53851 Lappeenranta
Finland

kalle.saastamoinen@lut.fi

Pasi Luukka
Laboratory of Applied Mathematics
Lappeenranta University of Technology
P.O. Box 20, FIN-53851 Lappeenranta
Finland

pasi.luukka@lut.fi

Abstract

In this paper we have derived four different fuzzy similarities from three extensions of t-norms mentioned in [1] pages 195-200. We have derived these similarities in the same spirit like they would have been used in Łukasiewicz algebra.

We will present four different similarities based on the extensions of t-norms these are known as $T_{\inf}$-extension, P-extension, E-extension and H_n extension. Similarity based on H_n extension is made by implementing different parameters into E-extension based similarity.

We will use these similarities in classification tasks and show some classification results of these similarities when they are used as classifiers. We will show that actually similarity relation based on $T_{\inf}$-extension is based on Łukasiewicz algebra and that it gives the best classification results from similarities derived and tested here. We will also show that weighting will make classification results better.

This study will be a short preliminary work as we aim to move on to study some other algebraic structures than Łukasiewicz. In our extended version of this paper we will investigate the use of some other means than arithmetic with these equivalence relations presented here.

Keywords: T-norm, Fuzzy classifier, Similarity measures

1 Introduction

Lately we have studied classifiers that are based on the use of t-norm called Łukasiewicz. We have gained some good results with this structure in the classification tasks and now we are going test some other possible t-norms that can be used in similar ways. In classification we are using the extensions of t-norms these are known as $T_{\inf}$-extension, P-extension, E-extension and H_n extension. In future we aim to study more the properties and the use of different t-norm based fuzzy similarities.

2 Comparison of Different Similarities Based on T-norms

In formulas below we are examining a choice situation where features of different objects have been normed in values between [0,1]. Let X be the set of m objects and the similarity values of the features between n objects that are known marked as $f_1, ..., f_n$. Since we know these similarity values, we can choose the object that has the highest total similarity value. Now we are trying to find for object x_i similar object x_j, where $1 \leq i, j \leq m$ and $i \neq j$. By choosing following t-norms for features of the objects we get n fuzzy similarities for comparing two objects (x_1, x_2):

Similarity relation based on $T_{\inf}$-extension:

$$S_{T_{\inf}}\langle x_1, x_2\rangle = \sum_{i=1}^{n} \frac{1}{n}(1 - |x_1(f_i) - x_2(f_i)|) \qquad (1)$$

Similarity relation based on P-extension:

$$S_P\langle x_1,x_2\rangle = \sum_{i=1}^{n}\frac{1}{n}(1-x_1(f_i)+x_1(f_i)x_2(f_i)) \\ (1-x_2(f_i)+x_2(f_i)x_1(f_i)) \quad (2)$$

Similarity relation based on E-extension:

$$S_E\langle x_1,x_2\rangle = \sum_{i=1}^{n}\frac{1}{n}(\frac{1+x_2(f_i)-x_1(f_i)}{1+x_2(f_i)-x_1(f_i)x_2(f_i)} \\ \frac{1+x_1(f_i)-x_2(f_i)}{1+x_1(f_i)-x_2(f_i)x_1(f_i)}) \quad (3)$$

One way to adjust similarity based on E-extension t-norm with additional parameter w is following, when we get a similarity relation called **Similarity relation based on H_n extension:**

$$S_H\langle x_1,x_2\rangle = \sum_{i=1}^{n}\frac{1}{n}(\frac{1+wx_2(f_i)-wx_1(f_i)}{1+wx_2(f_i)-wx_1(f_i)x_2(f_i)} \\ \frac{1+wx_1(f_i)-wx_2(f_i)}{1+wx_1(f_i)-wx_2(f_i)x_1(f_i)}) \quad (4)$$

3 Empirical Results

We used three different well known data sets which were taken from [2]. Data sets were splitted in half. One half was used for learning and the other half for testing.

Iris-data Set Data set consists of three classes and it is four dimensional.

Thyroid gland-data Set Data consist also of three classes and it is five dimensional.

Wine recognition-data Set Data consist of three classes and it is 13 dimensional.

In table 1 there is the classification results of the three similarities derived from t-norms. $T_{\inf}$ is actually the simple version of Łukasiewiczs algebra based similarity relation. As seen from the results $T_{\inf}$ gives better results with respect to P and E-extension similarity relations. In Figure 1 there is results from the H_n-extension similarity which is same as E-extension but with additional parameter w added. We studied the effects of this parameter with respect to classification results. As seen with proper choice of w results enhanced and this suggests that correct weighting might enhance results even further. Unfortunately we have so far been unable to find such parameters.

Table 1: Classification results (percentage of correct classifications) with different t-norms.

Data Set	$T_{\inf}$**-ext.**	**P-ext.**	**E-ext.**
Iris	94.67	72.00	72
Thyroid	89.81	87.04	83.33
Wine	95.51	89.33	82.02

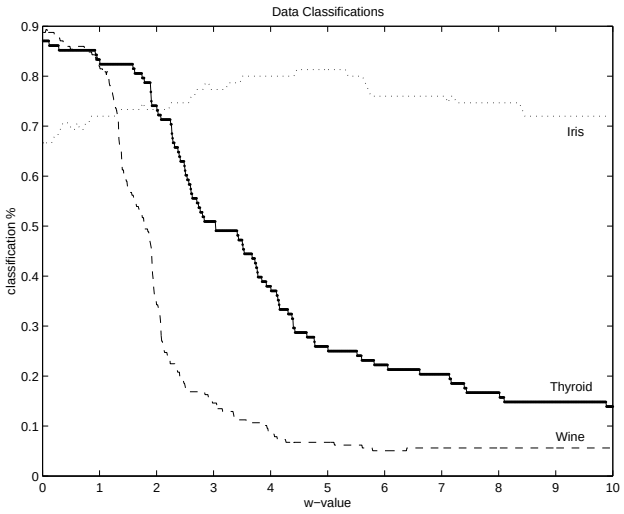

Figure 1: Results using similarity based on H_n extension with additional w-parameter with data sets.

4 Conclusions and Future

Similarity relation based on $T_{\inf}$-extension t-norm seems to work best compared to other similarities introduced here but complete study of other similarities derived from t-norms should be carried out. For future work will be left a thorough study of t-norms and how different metrics could be implemented in them.

References

[1] Lowen, R.: *Fuzzy Set Theory*. Kluwer Acad. Publishers, Dordrecht, 1996.

[2] UCI Repository of Machine Learning Databases network document. Referenced 4.11.1999. Available: *ftp://ftp.ics.uci.edu/pub/machine-learning-databases*