Iranian Journal of Fuzzy Systems

Volume 17, Number 5, (2020), pp. 137-146

University of Sistan and Baluchestan

On the use of Heronian means in a similarity classifier

O. Kurama[1]

[1] *Department of Mathematics, College of Natural Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda*

onesfolekurama@gmail.com, kurama@cns.mak.ac.ug

Abstract

This paper introduces new similarity classifiers using the Heronian mean, and the generalized Heronian mean operators. We examine the use of these operators at the aggregation step within the similarity classifier. The similarity classifier was earlier studied with other operators, in particular with an arithmetic mean, generalized mean, OWA operators, and many more. The two classifiers here are tested on four real world data sets, i.e., echocardiogram, fertility, horse-colic, and lung cancer. Three previously studied similarity classifiers are used as benchmarks to the new approaches. We observe that the similarity classifier with a generalized Heronian mean produces good classification results for the tested data sets, and is therefore more suitable for use in these classification problems.

Keywords: Similarity classifier, Heronian mean, aggregation, classification.

1 Introduction

Classification is one of the tasks that are central in data mining as well as data analysis. Data analysis is concerned with finding special structures within data sets, that can be helpful in guiding improvements in those areas of study. In particular, data analysis requires proper investigation, evaluation of the given data, and also involves drawing conclusions. Data analysis is usually burdened with uncertainties coming from variability, precision of sensors used in data collection, and sometimes vagueness from humans giving the data [1]. It is essential to develop methods that can cater for these uncertainties. At some stage, we need to identify the relationship of components in a data set, or sometimes map new samples to existing classes, here classifiers are required. Classifiers play an important role in data analysis, and several classification procedures have been suggested in literature, [5, 8, 21]. Classifiers also take a central role in many pattern recognition systems.

Similarity based classification methods were examined by Luukka *et al*. [21, 20]. In their work, a classifier applying the generalized mean in aggregating similarities was developed. Several classifiers based on computation of similarities have been developed and also tested on real world data sets, [10, 11, 12, 13, 19]. These models are constructed on the basis of fuzzy set theory, and use different aggregation operators in aggregating similarity vectors. The type of aggregation operator used affects the classification accuracy of a classifier. For more information on the theory of aggregation operators and some examples, see the detailed literature [2, 3, 6, 22]. In particular, this paper applies the Heronian mean and the generalized Heronian mean operators to aggregate similarity vectors. Initially, the Heronian mean occurred in an Egyptian manuscript dated 1850 B.C., [3], but it has been generalized to other forms by several researchers. Recently, applications of the Heronian mean operators in decision making have been examined in several papers [18, 16, 17, 26, 25]. Heronian means have also been applied in other areas, for more details, see [4, 14] and [24].

In this paper, we show that the use of the Heronian mean, and generalized Heronian aggregation operators in the similarity classifier, yields good classification results. This is a classification problem where we are interested in developing methods that will perform classification as accurate as possible. Two separate classification models are developed, and their performance tested on real world data sets. Experimental codes have been made with MATLAB^{TM} software in each case. The procedure starts with splitting the data set into two parts: the training and the testing sets. It is assumed that the training set has class labels, and our aim is to map the testing set samples into their possible

Corresponding Author: O. Kurama
Received: January 2019; Revised: October 2019; Accepted: April 2020.

classes. For each class in the training set, an ideal vector that best represents the class is computed. These ideal vectors are used for comparison purposes. The data sets used here are multi-attribute so that each sample (or object) in the data has numerical values under each attribute. This means that the objects we are classifying are n-dimensional vectors. The next step is to compute similarities between each object and each ideal vector. In this case we get n similarity values across all attributes for each class, which need some aggregation. Here aggregation is done using a Heronian mean in one case, and using the generalized Heronian mean in another case. An aggregated similarity value is obtained for each class in relation to the sample vector from the testing set. The sample then belongs to the class with which it has the highest similarity value. Details on this process are presented in section three.

Previously studied similarity classifiers applied the arithmetic mean, generalized mean, OWA operators, λ-averaging operator, Bonferroni mean operators, and others [20, 10, 11, 12, 13, 19]. These methods require weights, which were generated via linguistic quantifiers. However, the new approaches in this paper have applied the Heronian mean operators, which do not require weight generation, yet yield promising results. An added advantage of using the generalized Heronian mean is that, it takes into account the coalition of input arguments to be aggregated, making it suitable for practical applications [2].

The remainder of this paper is arranged as follows: Section two presents definitions and concepts required by the reader. In section three, new classification models are given. Numerical results are discussed in Section four, starting with applied data sets, and then to experimental results. The last section gives a brief discussion of key issues from new models compared to the benchmarks.

2 Preliminaries

Underlying definitions and key concepts are presented in this section. Heronian mean based aggregation operators are in particular mentioned before going into similarity measures. It should be noted that there are several aggregation operators in literature, but only those used in this paper are discussed.

Definition 2.1. (Heronian mean, [25]) *Let $(x_1, x_2, ..., x_n) \in [0, 1]$ be a set of non negative real numbers. The Heronian mean (HR) is a function:*

$$HR(x_1, x_2, ..., x_n) = \frac{2}{n(n+1)} \sum_{i=1}^{n} \sum_{j=i}^{n} \sqrt{x_i x_j}. \tag{1}$$

From Equation (1), it can be observed that when $n = 2$, the Heronian mean operator can be written in a basic form as: $HR(x_1, x_2) = \frac{1}{3}(2M + G)$, where $M = \frac{x_1+x_2}{2}$ is the arithmetic mean, and $G = \sqrt{x_1 x_2}$ is the geometric mean [2]. The Heronian mean is an aggregation operator since it satisfies the three basic properties of an aggregation operator, i.e it is monotonic, continuous, and satisfies boundary conditions on $[0, 1]$. [3, 26].

Yu and Wu [26], suggested a generalized form of the Heronian mean operator in Equation (1). In the general case, two parameters α and q were introduced, and these are essential in explaining interaction between attributes in the practical sense. Next is the definition of the generalized Heronian mean aggregation operator [25].

Definition 2.2. (Generalized Heronian mean) *Let $\alpha, q \geq 0$ be parameters that are not simultaneously zero, and $(x_1, x_2, ..., x_n) \in [0, 1]$ be a vector with at least one $x_i \neq 0 \ \forall \ i = 1, 2, ..., n$. The generalized Heronian mean (GHR) is the function,*

$$GHR^{\alpha,q}(x_1, x_2, ..., x_n) = \left(\frac{2}{n(n+1)} \sum_{i=1}^{n} \sum_{j=i}^{n} x_i^{\alpha} x_j^{q} \right)^{\frac{1}{\alpha+q}} \tag{2}$$

Notice that for $\alpha = p = 1/2$, Equation (2) reduces to Equation (1) without loss of generality, [25]. Other generalizations of the Heronian mean have been studied by Yu, [26], but in this article we have only applied two operators. Aggregation operators in equations (1) and (2) are applied in aggregating similarities within the similarity classifier, as explained in section 3.

The next concept applied in this article is similarity measures. In order to preserve the notion of classifiers in fuzzy set theory, here we concentrate on similarity measures in generalized Łukasiewicz-structure [21]. Two objects with several attributes can be compared effectively in the framework of Łukasiewicz-structure, without distorting similarities. Notice that the mean of many similarities is still a similarity when examined in the generalized Łukasiewicz-structure [23]. While it may sound trivial to determine similarities between objects, it is rather a complicated notion to realize in practice. We will briefly explain how this is done, having in mind that objects in question have measurable attributes that can be compared.

Consider two objects x_1 and x_2 in a fuzzy set X, with n attributes $(t_1, t_2, ..., t_n)$. The task is to compare the two objects across all attributes, and we recall the equivalence relation here such that: $x_1(i) \leftrightarrow x_2(i), \quad i = 1, 2, ..., n$. By choosing the Łukasiewicz-structure as the space for attributes, we obtain n fuzzy similarities S_{t_i} such that: $S_{t_i}\langle x_1, x_2\rangle = x_1(t_i) \leftrightarrow x_2(t_i)$. Thus, the fuzzy similarity is defined by: [20]

$$S_{t_i}\langle x_1, x_2\rangle = \frac{1}{n}\sum_{i=1}^{n}\left(x_1(t_i) \leftrightarrow x_2(t_i)\right). \tag{3}$$

Notice that Equation (3) holds only in Łukasiewicz-structure, and the equivalence relation $x_1 \leftrightarrow x_2$ can as well be defined by: $x_1 \leftrightarrow x_2 = 1 - |x_1 - x_2|$. In generalized Łukasiewicz-structure, this relation can be written as: $x_1 \leftrightarrow x_2 = 1 - |{x_1}^p - {x_2}^p|^{1/p}$, where p is a parameter for similarity measure [20]. We are now in position to define the similarity measure in the generalized Łukasiewicz-structure as:

$$S_{t_i}\langle x_1, x_2\rangle = \frac{1}{n}\sum_{i=1}^{n}\sqrt[p]{1 - |(x_1(t_i))^p - (x_2(t_i))^p|}. \tag{4}$$

Equation (4) clearly uses the arithmetic mean as a matter of choice, other means can also be applied. Using the generalized Heronian mean in place of arithmetic mean, Equation (5) is obtained.

$$S\langle x_1, x_2\rangle = \left[\frac{2}{n+1}\sum_{i=1}^{n}\left(1 - |(x_1(t_i))^p - (x_2(t_i))^p|\right)^{\frac{\alpha}{p}}\left(\frac{1}{n}\sum_{j=i}^{n}\left(1 - |(x_1(t_i))^p - (x_2(t_i))^p|\right)^{\frac{q}{p}}\right)\right]^{\frac{1}{\alpha+q}} \tag{5}$$

where α and q are parameters in the generalized Heronian mean operator, and p is the parameter for similarity measure. When $\alpha = q = 1/2$ are the settings in Equation (5), we get a similarity measure with the basic Heronian mean. In this case, we have Equation (6).

$$S\langle x_1, x_2\rangle = \frac{2}{n+1}\sum_{i=1}^{n}\left(1 - |(x_1(t_i))^p - (x_2(t_i))^p|\right)^{\frac{1}{2p}}\left(\frac{1}{n}\sum_{j=i}^{n}\left(1 - |(x_1(t_i))^p - (x_2(t_i))^p|\right)^{\frac{1}{2p}}\right) \tag{6}$$

We note that the larger the similarity value, the more similar the objects are. Similarity measure based classification with Heronian means applied in aggregation is explained in the next section.

3 Similarity classifier

Two different classification models are developed in this section. Here, we are interested in assigning classes to the testing part of the data set as accurate as possible. Essentially, the training part of the data set is properly labeled so that we can find a vector (ideal vector) that suitably represents each class. Objects to be classified are vectors with entries in the unit interval.

3.1 The classifier with a Heronian mean operator

Let $C_1, C_2, ..., C_K$ be the K possible classes in an n-dimensional attribute space. Each vector in the attribute space takes the form $\mathbf{x} = (x_1, x_2, ..., x_n)$. These vectors are scaled so that each entry of $\mathbf{x}$ is in the unit interval, $[0, 1]$. The process starts with computing ideal vectors from the training set for each class. For example, an ideal vector for class j takes the form: $\mathbf{v}_j = (v_j(1), v_j(2), ..., v_j(n))$, where $v_j(1)$ is the mean of entries under attribute t_1 in the j^{th} class. The similarity between a sample $\mathbf{x}$ from the testing set (to be classified) with ideal vectors can be determined here using Equation (7), which originates from Equation (6).

$$S\langle \mathbf{x}\ , \mathbf{v}\rangle = \frac{2}{n+1}\sum_{i=1}^{n}\left(1 - |(x(t_i))^p - (v(t_i))^p|\right)^{\frac{1}{2p}}\left(\frac{1}{n}\sum_{j=i}^{n}\left(1 - |(x(t_i))^p - (v(t_i))^p|\right)^{\frac{1}{2p}}\right) \tag{7}$$

where $\mathbf{x}, \mathbf{v} \in [0, 1]^n$, p is a parameter for similarity measure, and $(t_1, t_2, ..., t_n)$ are attributes.

The actual classification process in the testing part is described by the pseudo code in Algorithm 1. In this algorithm, similarities are aggregated using the Heronian mean. The choice of the class to which the test sample belongs is determined by the highest similarity value.

Algorithm 1 : Pseudo code for the similarity classifier with the Heronian mean operator.

Require: $data[1,\ldots,n], ideals,\ p$
  for $i = 1$ to n **do**
    for $j = 1$ to K **do**
      for $k = 1$ to n **do**
        $S(i,j,k) = [1 - |(data(i,k))^p - (ideal(j,k))^p|]^{1/p}$
      end for
    end for
  end for
  for $i = 1$ to n **do**
    for $j = 1$ to K **do**
      for $k = 1$ to n **do**
        $Sum1 = sum\left(S(i,:,k)^{1/2}\right)$
      end for
      $S(i,j) = \frac{2}{n(n+1)}\left(sum\left(S(i,:,k)^{1/2}.*Sum1\right)\right)$
    end for
  end for
  for $i = 1$ to n **do**
    $class(i) = find(S(i,:) == max(S(i,:)))$
  end for

3.2 The classifier with a generalized heronian mean operator

In this classification method a generalized Heronian mean operator is used in aggregating similarities. These similarities are obtained from test samples and ideal vectors as explained in subsection 3.1, using Equation (5). In this case, the similarity between a sample $\mathbf{x}$ and an ideal vector $\mathbf{v}$ is computed from:

$$S\langle \mathbf{x}\ , \mathbf{v}\rangle = \left[\frac{2}{n+1}\sum_{i=1}^{n}\left(1 - |(x(t_i))^p - (v(t_i))^p|\right)^{\frac{\alpha}{p}}\left(\frac{1}{n}\sum_{j=i}^{n}\left(1 - |(x(t_i))^p - (v(t_i))^p|\right)^{\frac{q}{p}}\right)\right]^{\frac{1}{\alpha+q}} \tag{8}$$

where α, q are parameters in the generalized Heronian mean operator, and p is a parameter for similarity measure. These parameters affect classification accuracy of the classifier, so they need to be tuned for optimal values.

The classification process applying the generalized Heronian mean operator in aggregation is described by the pseudo code in Algorithm 2. Ideal vectors are computed from the training set, and they are used for comparison purposes. The next step is to compute similarities between the sample to be classified and each ideal vector across all attributes. After this, each vector of similarity values is aggregated to get a single similarity value for each class. The sample x is classified in accordance with Equation (9) as follows:[13]

$$\text{Decide that } \mathbf{x} \in \mathrm{C_i} \text{ if } \ \mathrm{S}\langle \mathbf{x}\ , \mathbf{v}\rangle = \max_{\mathrm{i=1,2,\ldots,K}} \mathrm{S}\langle \mathbf{x}\ , \mathbf{v}\rangle. \tag{9}$$

This similarity classifier provides a partial membership for each class through the similarity concept. Memberships of objects are defined in Lukasiewicz structure. One of the underlying reasons why the Lukasiewicz structure is chosen is that, the mean of many fuzzy similarities is still a fuzzy similarity in this structure.

Algorithm 2 : Pseudo code for the similarity classifier with the generalized Heronian mean operator.

Require: $data[1,\ldots,n], ideals,\ p,\ \alpha,\ q$
  for $i = 1$ to n **do**
    for $j = 1$ to K **do**
      for $k = 1$ to n **do**
        $S(i,j,k) = [1 - |(data(i,k))^p - (ideal(j,k))^p|]^{\frac{1}{p}}$
      end for
    end for
  end for
  for $i = 1$ to n **do**
    for $j = 1$ to K **do**
      for $k = 1$ to n **do**
        $Sum1 = sum\left(S(i,:,k)^q\right)$
      end for
      $S(i,j) = \left[\frac{2}{n(n+1)}\left(sum\left(S(i,:,k)^{\alpha}.*Sum1\right)\right)\right]^{\frac{1}{\alpha+q}}$
    end for
  end for
  for $i = 1$ to n **do**
    $class(i) = find(S(i,:) == max(S(i,:)))$
  end for

4 Numerical analysis

4.1 A note on applied data sets

Data sets used in this paper were taken from UCI Machine Learning Repository [15]. Three medical data sets, i.e echocardiogram, fertility, and lung cancer have been used. Also horse-colic data set from the veterinary domain is used. Next we give brief descriptions of these data sets, including number of classes, number of attributes and instances present. Information in all applied data sets is summarized in Table 1.

The main aim of echocardiogram data set is to classify whether a patient will survive for at least one year following a heart attack. It has 12 measured attributes which include: 1. Survival. 2. Still alive. 3. Age at heart attack. 4. Pericardial effusion. 5. Fractional shortening. 6. E-point septal separation (epss). 7. Left ventricular end-diastolic dimension (lvdd). 8. Wall-motion-score. 9. Wall-motion-index. 10. Mult. 11. Name. 12. Alive at one year or not alive (class attribute).

Fertility data set was donated by David Gil et al. from Lucentia Research Group, Department of Computer Technology, University of Alicante - Spain [7]. The purpose is to determine whether a person is fertile or not. It was obtained by analyzing semen samples from 100 volunteers. It is believed that sperm concentration is related to a number of factors such as: environment, health status of a person, social demographic data, and life habits. All these factors were captured in measuring 10 attributes, which include: 1. Season (winter, spring, summer, fall). 2. Age at time of analysis (people between 18 and 36 years were considered). 3. Childish diseases (such as chicken pox, measles, mumps, and polio). 4. Accident or serious trauma. 5. Surgical intervention. 6. High fevers in the last year. 7. Frequency of alcohol consumption. 8. Smoking habit. 9. Number of hours. 10. Output (class attribute that shows whether the sample is normal or altered).

In horse-colic data set there are 386 samples and 28 attributes, these are: 1. Surgery, 2. Age, 3. Hospital Number, 4. Rectal temperature, 5. Pulse, 6. Respiratory rate, 7. Temperature of extremities, 8. Peripheral pulse, 9. Mucous membranes, 10. Capillary refill time, 11. Pain, 12. Peristalsis, 13. Abdominal distension, 14. Nasogastric tube, 15. Nasogastric reflux, 16. Nasogastric reflux PH, 17. Rectal examination - feces, 18. Abdomen, 19. Packed cell volume, 20. Total protein, 21. Abdominocentesis appearance, 22. Abdomcentesis total protein, 23. Outcome, 24. Surgical lesion, 25, 26, 27 are type of lesion, and 28. Cp data. Most of these variables are not significant, and only 11 attributes were considered [15].

The last data set used is lung cancer which has three classes, 56 attributes, and 32 instances. It indicates that there are a number of missing values, which makes this data set difficult to classify. The author of this data set does not give much information, but passed usage shows that Hong and Young obtained an accuracy of 77% using the KNN method [9].

Data set name	Number of classes	Number of attributes	Number of instances	Domain
Echocardiogram	2	12	132	Medical
Fertility	2	10	100	Medical
Horse-colic	2	28	368	Veterinary
Lung cancer	3	56	32	Medical

Table 1: Properties of applied data sets.

4.2 Experimental results

Experiments are arranged so that each data set is divided into two equal parts, one part for training and the other for testing. In each experiment, divisions are randomly done 30 times into the training and testing, and the resulting mean accuracies recorded. Mean classification accuracies attained and their corresponding variances, with new methods are presented in Tables 2 and 3. In the similarity classifier with a generalized mean operator, there are three parameters of concern: p for similarity measure, and α, q in the generalized Heronian mean aggregation operator. To study all the three parameters with respect to classification accuracy creates a 4-dimensional problem which may not easily be visualized geometrically. In order to investigate this properly, we fix parameter q in the first experimental run and study the rest with respect classification accuracy. In other runs, parameter q is changed in the direction of increasing accuracy until there is no more improvement. The value of q at which maximum accuracy is obtained is then recorded as the optimal value of q.

Data set	Mean accuracy %	Variance
Echocardiogram	91.30	$6.6733 * 10^{-4}$
Fertility	75.47	0.0078
Horse-colic	82.93	$4.6378 * 10^{-4}$
Lung cancer	82.94	0.0049

Table 2: Classification results using Heronian mean operator

Data set	Parameter, q	Mean accuracy %	Variance
Echocardiogram	0.5	91.92	$6.8814 * 10^{-4}$
Fertility	12	87.20	$5.1310 * 10^{-4}$
Horse-colic	5	84.32	$6.2755 * 10^{-4}$
Lung cancer	5	83.33	0.0046

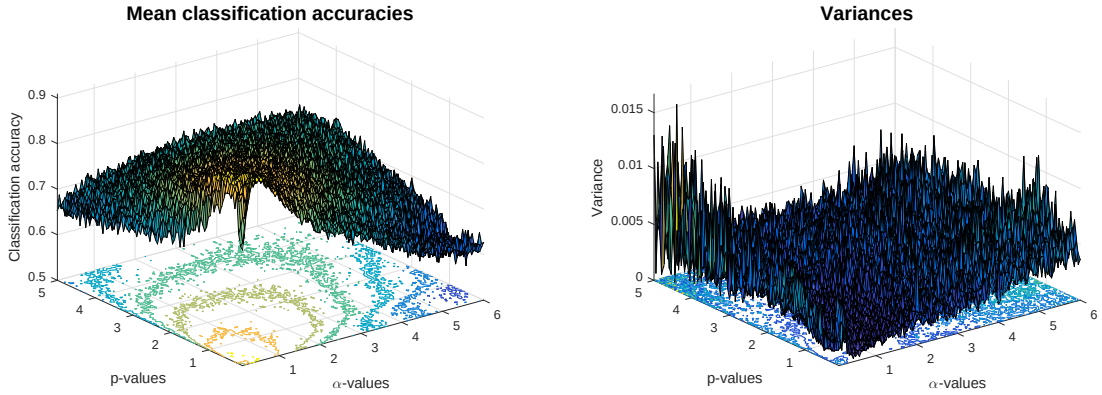
Table 3: Classification results using generalized Heronian mean operator

Tables 4 - 7 show mean classification accuracies from new classifiers and their benchmarks, with corresponding variances. The highest mean classification accuracy in each table is indicated with bold case. Three-dimensional plots corresponding to the highest mean classification accuracies are given in Figures 1 - 4. Results achieved in each data set are presented next.

Using echocardiogram data set the highest mean classification accuracy of 91.92% was attained with the method that applies a generalized Heronian mean operator in aggregation. The rest of other mean accuracies from benchmark classifiers are well below the mean accuracies obtained from the new methods. Results are reported in Table 4 including variances along side each mean accuracy. In Figure 1, variation of mean classification accuracy with parameters p and α is plotted for a fixed value of q. After proper optimization, the highest mean classification is achieved when q is set at 0.5.

Similarity classifier with:	Mean accuracy %	Variance
Arithmetic mean	86.89	$8.9022 * 10^{-4}$
Generalized mean	90.11	$6.1582 * 10^{-4}$
OWA operator	90.82	$5.3512 * 10^{-4}$
Heronian mean	91.30	$6.6733 * 10^{-4}$
Generalized Heronian mean	**91.92**	$6.8814 * 10^{-4}$

Table 4: Classification results with echocardiogram data set.



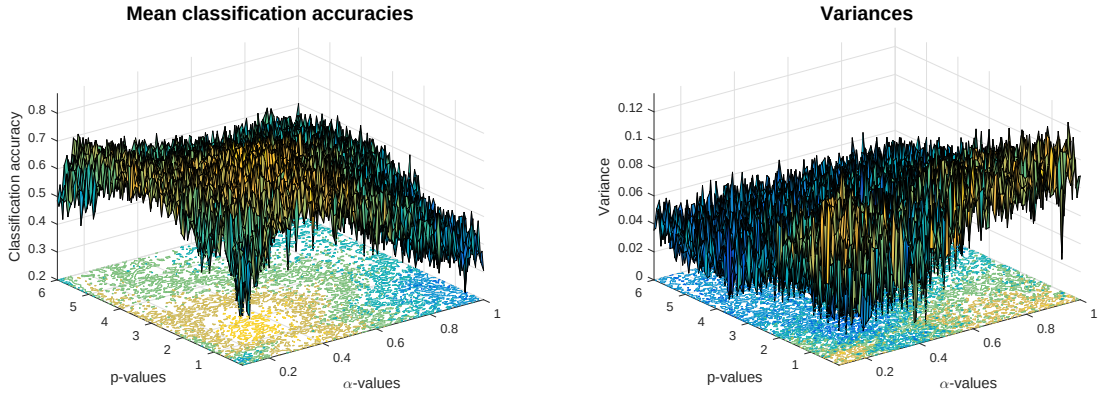
(a) (b)

Figure 1: Mean classification accuracies (a) and the variances (b) obtained from echocardiogram data set with generalized Heronian mean at q = 0.5.

The similarity classifier with the generalized Heronian mean achieved a mean classification accuracy of 87% with fertility data set. This was the highest possible mean accuracy compared to others. The classifier with a Heronian mean was the second best and the rest followed in order from OWA operator to Arithmetic mean classifiers as the worst. Notice that in this case, the highest mean accuracy was got when parameter q is fixed at 12. More detailed results can be seen in Table 5. Also a three-dimensional plot that investigates the changes in accuracy with respect to other parameters is shown in Figure 2.

Similarity classifier with:	Mean accuracy %	Variance
Arithmetic mean	66.87	0.0031
Generalized mean	69.07	0.0073
OWA operator	74.33	0.0085
Heronian mean	75.47	0.0078
Generalized Heronian mean	**87.20**	$5.1310 * 10^{-4}$

Table 5: Classification results with fertility data set.



(a) (b)

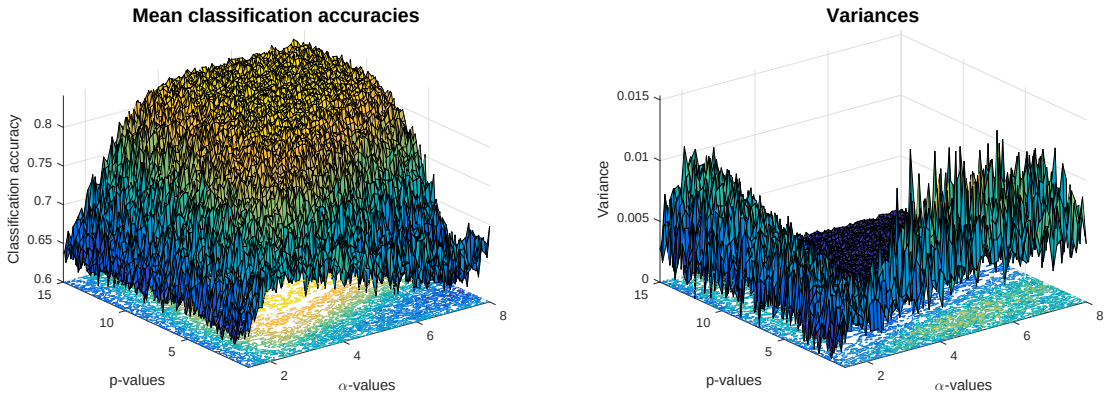
Figure 2: Mean classification accuracies (a) and the variances (b) obtained from fertility data set with generalized Heronian mean at q = 12.

With horse-colic data set, the similarity classifier with the generalized Heronian mean achieved the highest mean classification accuracy of 84.32%. Previously, the classifier with an OWA operator had achieved a mean accuracy of 82.18% as the best. Clearly the new classifiers are better than the benchmarks as seen from Table 6. Notice that variances corresponding to mean classification accuracies are too small, which indicates that these mean accuracies are quite close to the true mean. In this experiment, the optimal value of q was 5. It can also be seen from Figure 3 that other parameters p and α can be set around $p = 8$, and $\alpha = 5$ to achieve optimal performance of this version of the similarity classifier.

Similarity classifier with:	Mean accuracy %	Variance
Arithmetic mean	78.90	$5.7930 * 10^{-4}$
Generalized mean	80.20	$4.3550 * 10^{-4}$
OWA operator	82.18	$8.1788 * 10^{-4}$
Heronian mean	82.93	$4.6378 * 10^{-4}$
Generalized Heronian mean	**84.32**	$6.2755 * 10^{-4}$

Table 6: Classification results with horse-colic data set.

The new similarity classifier achieved a mean classification accuracy of 83.33% with a variance of 0.0046. The classifier with a Heronian mean achieved an accuracy of 82.94% which is quite close to the highest accuracy. Detailed results from other methods can be seen in Table 7. Also with q fixed at 5, a three dimensional plot of mean accuracy



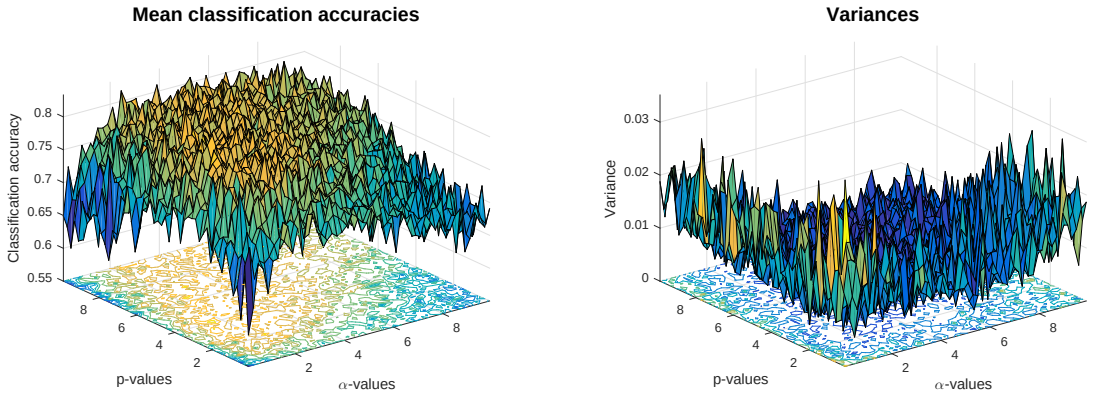
(a) (b)

Figure 3: Mean classification accuracies (a) and the variances (b) obtained from horse-colic data set with generalized Heronian mean at q = 5.

with other parameters is shown in Figure 4. From this figure, other parameters can approximately be chosen around $p = 6$ and $\alpha = 4$.

Similarity classifier with:	Mean accuracy %	Variance
Arithmetic mean	80.59	0.0055
Generalized mean	82.16	0.0032
OWA operator	81.96	0.0043
Heronian mean	82.94	0.0049
Generalized Heronian mean	**83.33**	0.0046

Table 7: Classification results with lung cancer data set.



(a) (b)

Figure 4: Mean classification accuracies (a) and the variances (b) obtained from lung cancer data set with generalized Heronian mean at q = 5.

5 Discussion

Two classification models that utilize Heronian aggregation operators have been presented. The operators were applied during the aggregation step within the similarity classifier. Three benchmark similarity classifiers have been considered here, which use, the arithmetic mean, the generalized mean, and the OWA operators. Results achieved with the new

methods look better than those achieved earlier in the benchmark classifiers. In these classification problems, even a small improvement in classification accuracy is crucial. For example in the medical field where most of our data sets are taken, a small improvement can save several human lives. Besides better classification accuracies achieved with the two new classifiers here, these methods also do not need any generation of weights as was needed in the benchmark methods. A comparison in performance of the new methods and the other classifiers follows: The method using the generalized Heronian mean operator achieved highest mean classification accuracies in all the four tested data sets. It can be observed that the method with the Heronian mean achieved second best results in all the tested data sets. Taking the similarity classifier with an arithmetic mean as the standard classifier, it can be viewed from results, that there are improvements in all achieved classification accuracies. There is an improvement of 5.03% achieved with the new method with echocardiogram data set compared with the standard similarity classifier. In the same manner, an improvement of 20.33% was achieved with fertility data set. With Horse-colic data set, an improvement of 5.42% was obtained. Lastly, the new method yielded an improvement of 2.74% with Lung cancer data set. Since these data sets are medical related, even a small improvement should be appreciated. It is worth to mention that different classification accuracies were obtained from different data sets. This implies that these new classification methods are data dependent. In conclusion, we need to always test the new methods on each data set in order to establish the most suitable model.

Acknowledgements

The author would like to thank the anonymous reviewers for their constructive comments that have helped to improve this paper.

References

[1] H. Bandemer, W. Näther, *Fuzzy data analysis*, Kluwer Academic Publishers Group, Dordrecht, The Netherlands, 1992.

[2] G. Beliakov, A. Pradera, T. Calvo, *Aggregation functions: Aguide for practitioners*, Springer-Verlag, Berlin, 2007.

[3] P. S. Bullen, *Handbook of means and their inequalities*, Kluwer Academic Publishers, Dordrecht, 2003.

[4] A. Cizmesija, *A new sharp double inequality for generalized Heronian, harmonic and power means*, Computers and Mathematics with Applications, **64** (2012), 664-671.

[5] E. Diday, *Classification and multivariate analysis for complex data structures*, Springer-Verlag Berlin, 2011.

[6] D. Dubois, H. Prade, *On the use of aggregation aperations in information fusion processes*, Fuzzy Sets and Systems, **142** (2004), 143-161.

[7] D. Gil, L. J. Girela, D. J. Juan, J. M. Gomez Torres, M. Johnsson, *Predicting seminal quality with artificial intelligence methods*, Expert Systems with Applications, **39**(16) (2012), 12564-12573.

[8] A. D. Gordon, *Classification, second edition*, Chapman and Hall/CRC press, 1999.

[9] Z. Q. Hong, J. Y. Yang, *Optimal discriminant plane for a small number of samples and design method of classifier on the plane*, Pattern Recognition, **24**(4) (1991), 317-324.

[10] O. Kurama, P. Luukka, M. Collan, *An n-ary λ-averaging based similarity classifier*, International Journal of Applied Mathematics and Computer Sciences, **26**(2) (2016), 407-421.

[11] O. Kurama, P. Luukka, M. Collan, *A similarity classifier with Bonferroni mean operators*, Advances in Fuzzy Systems, Article ID 7173054, (2016), 1-11.

[12] O. Kurama, P. Luukka, M. Collan, *On a weighted ordered weighted averaging based similarity classifier*, Proceedings of the 17th IEEE International Symposium on Computational Intelligence and Informatics (CINTI 2016), Óbuda University, Budapest-Hungary, November, 2016.

[13] O. Kurama, *Similarity based classification methods with different aggregation operators*, Doctoral Dissertation, Lappeenranta University of Technology, 2017.

[14] V. Labunetsa, E. Osthaimerb, *Systematic approach to nonlinear filtering associated with aggregation operators: Part 1 SISO-filters*, Procedia Engineering, **201** (2017), 372-384.

[15] M. Lichman, *UCI machine learning repository, Irvine, CA: University of California*, School of Information and Computer Science, 2013.

[16] P. Liu, *Multiple attribute group decision making method based on interval-valued intuitionistic fuzzy power Heronian aggregation operators*, Computers and Industrial Engineering, **108** (2017), 199-212.

[17] P. Liu, J. Liu, J. M. Merigo, *Partitioned Heronian means based on linguistic intuitionistic fuzzy numbers for dealing with multi-attribute group decision making*, Applied Soft Computing, **64** (2018), 395-422.

[18] P. Liu, Z. Liu, Z. Xin, *Some intuitionistic uncertain linguistic Heronian mean operators and their application to group decision making*, Applied Mathematics and Computation, **230** (2014), 570-586.

[19] P. Luukka, O. Kurama, *Similarity classifier with ordered weighted averaging operators*, Expert Systems with Applications, **40**(4) (2013), 995-1002.

[20] P. Luukka, T. Leppälampi, *Similarity classifier with generalized mean applied to medical data*, Computers in Biology and Medicine, **36** (2006), 1026-1040.

[21] P. Luukka, K. Saastamoinen, V. Könönen, *A classifier based on the maximal fuzzy similarity in the generalized Łukasiewicz-structure*, Proceedings of the FUZZ-IEEE Conference, Melbourne, Australia, 2001.

[22] V. Torra, Y. Narukawa, *Modeling decisions, information fusion and aggregation operators*, Springer-Verlag, Berlin, Heidelberg, 2007.

[23] E. Turunen, *Mathematics behind fuzzy logic, advances in soft computing*, Physica-Verlag, Heidelberg, 1999.

[24] D. J. Yu, *Intuitionistic fuzzy geometric Heronian mean aggregation operators*, Applied Soft Computing, **13** (2013), 1235-1246.

[25] D. J. Yu, *Hesitant fuzzy multi-criteria decision making methods based on Heronian mean*, Technological and Economic Development of Economy, **23**(2) (2017), 296-315.

[26] D. J. Yu, Y. Y. Wu, *Interval-valued intuitionistic fuzzy Heronian mean operators and their applications to multi-criteria decision making*, African Journal of Business and Management, **6**(11) (2012), 4158-4168.