Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 82 (2016) 4 – 11

Procedia Computer Science

Symposium on Data Mining Applications, SDMA2016, 30 March 2016, Riyadh, Saudi Arabia

Fuzzy-Genetic Model for the Identification of Falls Risk Gait

Md Rafiul Hassan*[a], S M Arafat[a] and Rezaul K Begg[b]

[a]Dept of ICS, King Fahd University of Petroleum and Minerals, Dhahran 31261, KSA
[b]Dept of Biomechanics, Victoria University, VIC 8001, Australia

Abstract

This paper investigates the effectiveness of a hybrid genetic and fuzzy set algorithm for the recognition of gait patterns in falls risk patients. In a previous work, we have shown the usefulness of fuzzy set techniques for gait pattern identification. In this paper, we apply a genetic algorithm in conjunction with fuzzy logic rules to better select the optimal combination of pathological gait features for improved gait diagnostic capability. Gait features were calculated using minimum foot clearance data collected during continuous walking on a treadmill for 20 older adults. The subjects are composed of two groups, 10 individuals with normal gait, and 10 with a history of falls. Fuzzy rules were extracted from the gait dataset using subtractive clustering. The genetic algorithm was introduced in order to select the optimum combination of gait features. Using cross validation test data, the results indicated that the generalization performance, in terms of accuracy, for the hybrid system was 97.5%, compared to89.3% that was obtained using only the fuzzy system. The generalization performance of the gait classifier was also analyzed by determining the areas under the receiver operating characteristic plot. We observed that an improved gait classification performance became evident when the fuzzy system classifier used a small number of features that were selected by the genetic algorithm.

Peer-review under responsibility of the Organizing Committee of SDMA2016

Keywords: gait, falls risk; feature selection; genetic algorithms; fuzzy systems

1. Introduction

It is widely understood that the balance control mechanisms and the associated gait functions of the human locomotor system deteriorate with age. Some of these gait pattern changes can cause high health risks and can lead

* Corresponding author. Tel.: +966138607330.
E-mail address: mrhassan@kfupm.edu.sa

Peer-review under responsibility of the Organizing Committee of SDMA2016
doi:10.1016/j.procs.2016.04.002

to falls and injuries during movement. Falls during walking in older adults are a major social and health concern [1]. Also, falls risk have a close connection to gait degeneration and to altered gait balance that is due to ageing. Among the various types of falls, tripping and slipping account for the majority (i.e., more than 50%) of all fall cases [2].

Researchers have investigated potential parameters associated with tripping and falling risk of the elderly [3]. Such parameters include foot to ground clearance height during leg swing phase [4], minimum foot clearance [5], and lower limb joint angles [6].

One important parameter during the foot's flight (swing phase of the gait cycle) is the toe clearance, which is regarded as an important characteristic because it assists in adapting to uneven surfaces as well as avoiding obstacles. The toe comes very close to the ground during the mid-swing phase leaving a very small margin for error [7]. The minimum foot/toe clearance (MFC) during this event has been identified as the most critical factor that is associated with avoiding tripping, mainly because either a reduced magnitude or a highly varying MFC could lead to the increased risk of foot contact with obstacles [7].

One technique that could be applied for minimizing falls in older adults is to identify potential fallers through their modified gait characteristics, thereby engaging in intervention procedures to improve their gait function. Moreover, an objective model is necessary to link MFC information with falls risk individuals in an effort to determine MFC features that could be used to diagnose potential falls risk victims.

Due to the large intra-subject variation in gait [7], the task of identifying gait features that map potential fall risk patients is complex. One possible method is to develop nonlinear models based on computational intelligence approaches, in a bid to find relationships between MFC features and different falls risk categories (i.e., high risk or balance impaired individuals, low-risk or normal individuals with no-falls history, etc.). Such models can potentially have many applications to a number of fields including gait diagnostics, rehabilitation, assessment of at-risk gait, and other related areas.

In recent years, fuzzy inference models have emerged as powerful tools for solving many classification problems. Prior studies that applied fuzzy rules to classify gait types have shown success in detecting gait events, gait measurement [8], and classifying normal and ankle arthrodesis gait patterns [9]. In a previous investigation [10], we have demonstrated the effectiveness of fuzzy logic in classifying both healthy and falls risk gait patterns. In particular, we found out that subtractive clustering [11] for fuzzy rules generation using training data was quite useful and the trained fuzzy system was able to recognize the two gait types with 89% accuracy rate. In this paper, we investigate the use of a hybrid fuzzy genetic model in order to improve gait recognition. We have designed and implemented a genetic algorithm (GA) to select the optimal input feature combination for the fuzzy classification system. The objective of this hybridization was to achieve an improved accuracy rate in classifying at-risk gait in older populations. The performance of the hybrid GA fuzzy model was evaluated using accuracy rates as well as the receiver operating characteristics (ROC) curve. The remainder of this paper is organized in the following way: Section II offers an overview of the fuzzy genetic algorithm; Section III elaborates on the data, experiments, and results, and finally, Section IV concludes with a discussion and conclusion.

2. Overview of the model

A combination of GA and fuzzy logic system is proposed for gait pattern recognition and classification tasks. We applied subtractive clustering that was proposed by Chiu [11] in order to extract fuzzy rules from the dataset. GA was applied in parallel in order to select an optimal combination of input features that maximizes classification performance for both normal and falls risk gait subjects. The objective function used in the hybrid model maximizes classification accuracy. GA randomly selects input data features that are used to generate fuzzy rules. Next, the model predicts test cases that are evaluated using classification accuracies. Intuitively, GA selects the next generation that includes a combination of the well fitted input data features. Next, fuzzy rules are generated for the newly selected input data space for further testing. Thus the model includes three major steps:

1. Division of the training data into a number of clusters;
2. Generation of fuzzy rules for each of the clusters (as shown in Fig. 1);
3. Application of GA to select the best feature combination.

```
Algorithm: Generate fuzzy rules using subtractive clustering
Input: D: the training dataset, r: cluster radius
Output: Rule: the set of fuzzy rules
Cluster ← SubtractiveCluster(D,r)
Rule ← ∅
RuleIndex ← 0;
while Clusters= ∅ do
    for each features in data that belong to the respective cluster do
        RuleIndex ← RuleIndex + 1;
        Rule[RuleIndex].Mean ← calculate mean of the feature considering the clustered data;
    end
    Rule[RuleIndex].Coeff ← Obtain Coefficients using Least Square method and the data
     belonging the respective cluster with labels for this rule;
end
Return Rule;
```

Fig. 1 An algorithm for generating fuzzy rules using clusters that are formed by subtractive clustering.

2.1. Input data space division: Subtractive clustering

Before generating the fuzzy model, the input data space will be divided into a number of groupings. Each of the groups is then used to form individual fuzzy rules. In this model, we used subtractive clustering to divide the input data space, due to its efficiency in grouping data, as well as fast estimation of data clusters [11]. In subtractive clustering, the assumption is that every data point in the dataset is a potential cluster center, and that the potential has a numeric value that is computed using atonal '*n*' data points, according to the following equation:

$$\psi_i = \sum_{k=1}^{n} e^{-c\|x_i - x_k\|^2}$$

where, ψ_i is a measure of potential for each data point x_i to serve as cluster center, c= $4/r_a^2$, r_a is a positive constant that is a normalized radius defining the neighborhood, and $\|.\|$ is the Euclidean distance.

Using this technique, the data point that is associated with the maximum potential is selected as the first cluster center. Next, we determine subsequent cluster centers as follows. Suppose that the data point x_{c1} computes a maximum potential value ψ_{c1} , and therefore ψ_{c1} is the center of the first cluster, then, the potential for all the data in the group can be recalculated (i.e., adjusted) using the following equation:

$$\psi_i \Leftarrow \psi_i - \psi_{c1} e^{-\frac{4}{(1.25 r_a)^2}\|x_i - x_{c1}\|^2}$$

The above equation subtracts a certain amount from the 'potential' of each data point resulting in lesser 'potential' values for data points that are close to the first cluster center (x_{c1}). Therefore, those points that are close to the first cluster center will have very little likelihood of being selected as the next cluster center. The data point with the second maximum potential is subsequently selected as the second cluster center and the potential of the data points are again recalculated. This iterative process is repeated for the determination of all other cluster centers.

2.2. Fuzzy Logic

As outlined in Section II, above, data is grouped into *n* clusters. The next step generates fuzzy rules for every cluster. A fuzzy membership function is generated for every feature, per cluster. For example, consider a feature 'x'. The membership function '*M(x),'* or the degree to which x belongs to a specific cluster with a cluster center x_c and cluster standard deviation σ is:

$$M(x) = e^{-\frac{1}{2}\left(\frac{x-x_c}{\sigma}\right)^2} \tag{1}$$

We chose a Gaussian membership function mainly due to the simplicity of parameter fine tuning at the later, neural network training stage. Once the membership functions of each of the data features for the '*n*' clusters are defined, the i_{th} fuzzy rule, Y_i, is computed such that:

If x_1 is M_{i1} and x_2 is M_{i2} and … and x_j is M_{ij} then output is Y_i

The parameters of the initially generated fuzzy rules, x_c and σ , are fine-tuned using the back propagation-based gradient descent.

2.3. Selection of optimum features: Application of GA

In Section 2.2, above, all the attributes in the input data space are taken into consideration in the process of fuzzy rules generation. However, not all input features necessarily have significant influence on data classification. It is likely to have an improved classification performance if we provide only important features to the classifier. GA has been used in a number of studies to search for important features in the dataset. The survival of the fittest criterion of GA uses an adaptation mechanism that makes it significantly different from other search techniques. GA operates by using certain operators: selection, crossover and mutation. There should be an initial population to work with the three operators. The population consists of a set of strings that have their own fitness value. The fitness value is used to determine the probability for each string in the population of current generation to select the parent string from the current population. The crossover operator produces a child offspring from two selected strings by crossing over at some point. Using a time interval, mutation introduces variation to the population by changing a few bit values in the string. These processes are executed iteratively and the new next generations are repeatedly introduced until a stopping criterion is reached.

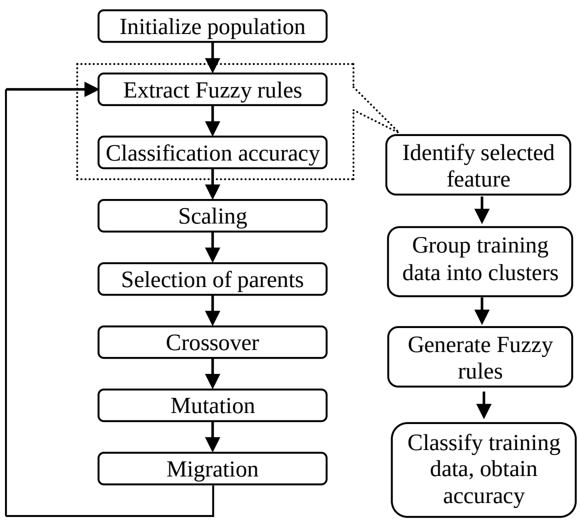

Fig. 2. GA process to obtain the optimal feature combination for fuzzy classifier

For our gait behavior classification problem we employ GA to search for the best combination of input features that are important for gait classification. Hence, the fitness value we have chosen is used either to maximize the classification accuracy or to minimize the classification error. Initially a set of strings (i.e., a population) is generated randomly. The objective function runs the classifier using the combination of features represented in a string and calculates the classification accuracy of features. The classification accuracy is set as the fitness value, and it is used to select a parent string in order to evaluate the next population.

Next, a crossover is operated to obtain a new set of strings from the parent strings. The mutation operation is applied to this process at the instruction given by the user.

For gait classification the string length is chosen at9 bits. Each bit represents a feature. For example, if the value of a bit at location ‘j’ in the string is ‘1’ then this indicates that the j^{th} feature is selected.

3. Experimental result

3.1. Participants

A gait dataset for developing and evaluating the hybrid and fuzzy models were taken from the Biomechanics Unit database at Victoria University, Australia. This set included the minimum foot clearance (MFC) of twenty older (>65 years old) adults. Age-wise, the set included 10 younger and normal adults, and 10 older adults with balance impairments. The subjects had no known injuries or natural abnormalities that would affect their normal gait.

3.2. Minimum Foot Clearance Features

The procedure for MFC data collection involves recording foot trajectory using the PEAK MOTUS 2D (Peak Technologies Inc., USA) motion analysis system while the participants walked continuously between 10 to 20 minutes at a self-selected comfortable walking speed. For more details you may refer to [4]. The raw data was digitally filtered using an optimal cut-off frequency and a Butterworth filter. Marker positions and shoe dimensions were used to predict the location of the shoe/foot end-point, i.e., the position on the shoe traveling closest to the ground at the time when MFC occurs using a 2D geometric model of the foot [12]. MFC was calculated by subtracting ground reference from the minimum vertical coordinate during the swing phase for each gait cycle.

Each subject’s MFC data were plotted as histograms showing individual MFC data and their respective frequencies. Features describing major statistical characteristics of these distributions were extracted such that they included: the minimum (MN), maximum (MX), mean, median (Q2), 25^{th}quartile (Q1), 75^{th} quartile (Q3), standard deviation (SD), skew (S) and kurtosis (K) [12].

3.3. Test strategy and performance metric

Gait features were normalized between -1 to +1 using their maximum and minimum values before being applied to the classifier. A five-fold cross validation was applied to generalize the overall performance of the method. To evaluate the classifier performance, measures of accuracy, ROC area (ROC_{area}), sensitivity and specificity were computed[12].

3.4. Results

The parameter values used for the GA model in this experiment are shown in Table 2, below. While extracting the fuzzy rules from the input space, we chose $r_a = 0.05$, following our previous study’s experimentation [10]. As stated earlier, the fitness value of GA is the classification accuracy and the objective function maximizes the classification accuracy. The hybrid GA-Fuzzy algorithm achieved the best performance for the following combination of input features out of the total nine features used in our previous study [10]: Q1, Q3, mean, SD and MX.

TABLE 1
PARAMETERS OF GENETIC ALGORITHM USED TO SETUP THE EXPERIMENTS

Population Type	Binary
Population Size	20
Elite Count	2
Crossover Fraction	0.8000
Migration Fraction	0.2000
Generations	100
Fitness Limit	100%
Initial Population	Random
Fitness Scaling	Rank Scaling

Table 2 illustrates overall test results for both models, i.e., GA-Fuzzy and fuzzy. The accuracy rate improved to 97.5%, and this is mainly due to the hybridization of the fuzzy classifier.

TABLE 2
RESULTS FOR CROSS-VALIDATED TEST DATA

Performance Metric	Fuzzy Model	Hybrid GA-Fuzzy Model
Accuracy	89.3%	97.5%
Sensitivity	0.87	0.975
Specificity	0.92	0.975
ROC_{area}	0.93	0.976

ROC results are depicted in Fig. 3 and they indicate a performance improvement of both sensitivity and specificity results. ROC_{area} was very close to 1,and this demonstrates an improved overall performance.

3.5. Representation of gait patterns in linguistic form

The output of the GA-Fuzzy classifier is a scalar numeric value. It would be useful if we translate specific gait pattern data into linguistic terms that would represent an overall quality of gait. In order to calculate these terms, we

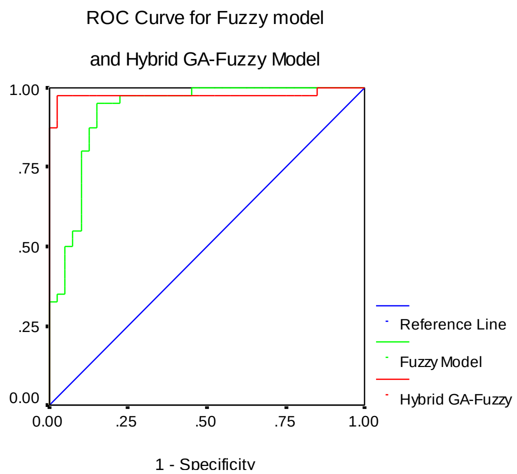

Fig. 3. ROC curve compares fuzzy and hybrid GA-Fuzzy classifiers.

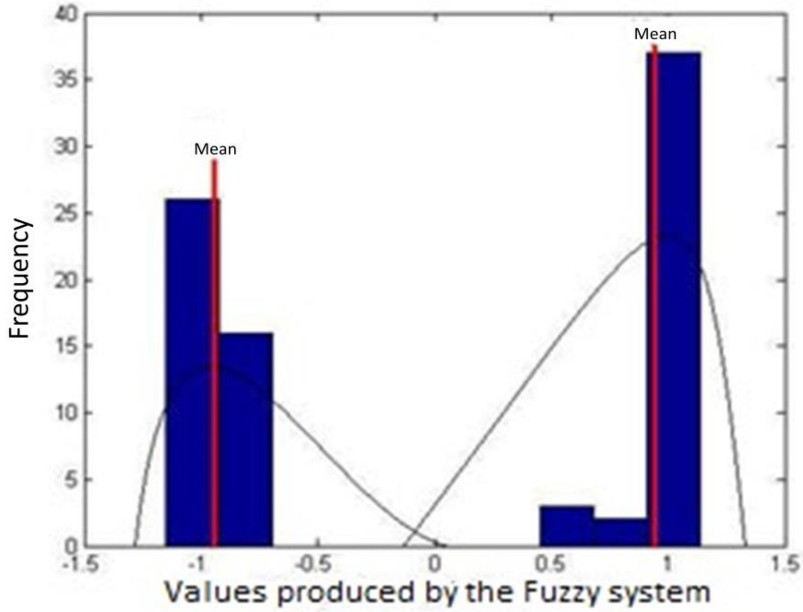

Fig. 4. Data distribution of classifier's output for two classes

compute the distribution of the predicted output that was produced by the GA-Fuzzy classifier.

We derived two distributions such that they represent the two gait classes (Fig.3). Subsequently, for any predicted output, the distance between a particular output and the two distributions can be calculated using the following equation [8]:

$$d = 1 - \frac{abs(y_{pred} - \mu_{pos})}{abs(\mu_{pos} - \mu_{neg})} \tag{2}$$

This distance 'd' quantifies the gait functional status, in the sense that a higher positive distance ('d') indicates that the gait pattern belongs to the positive class (i.e., normal), while a lower 'd' value indicates the gait pattern is more negative (i.e., pathological). In our experiments, we divided the distance ranges between 0 and 1 into five segments representing various gait functional qualities, as shown in Table 3. Outputs from the cross validated test results were grouped into these categories as per their equivalent distance measures. Table 3shows relative proportions of the two groups falling into the various categories, and Table 3 shows results for cross validated test data.

TABLE 3
GROUPING OF SUBJECTS USING CLASSIFIER OUTPUT

Gait Quality	Percentage (Out of 20 subjects)	Fuzzy Numeric Output Range
No Risk	45	0.76 to 1.1
Normal Risk	5	0.6 to 0.75
Mild Risk	0	0.26 to 0.5
High Risk	10	0.1 to 0.25
Very High Risk	40	-0.1 to 0.0

4. Discussion and Conclusion

Features play a vital role in every classification/detection problem. Features that have a significant role for one classification problem may not prove to be equally significant for another problem. Alternatively, a combination of significant features demonstrates its significance when used by a specific classifier. In this study, while computing the important features for the fuzzy classifier we noticed that the 'skew' feature may identify gait behavior with a very good accuracy. However, a combination of this variable with other variables was not found useful in improving the classification accuracy.

We applied GA in this study in order to identify the features that are most effective in gait classification. It revealed that a combination of the following five features would be most significant for the fuzzy classifier to detect falls-risk gait behavior:

Q1, Q3, mean, SD, MX

Table I shows the improvement of classification accuracy (both in terms of percentages and ROC area) of the GA-Fuzzy model over the individual fuzzy classifier (used in [9]).

The segmentation of the hybrid GA-Fuzzy classifier output helps us identify the status of gait patterns, precisely. This qualification is done by calculating the distance between the distributions of normal and abnormal gait and it revealed understandable and justifiable information about the status of gait patterns of individual subjects. Such qualification helps physicians and patients understand the human interpretable status about a subject's health.

Overall, the results suggest that significant improvement can be achieved through the use of GA-based feature selection, while classifying normal versus balance-impaired gait.

References

[1] Banerjee, T.; Keller, J.M.; Skubic, M.; Stone, E., "Day or Night Activity Recognition from Video Using Fuzzy Clustering," *IEEE Transactions on Fuzzy Systems*, 2014, Vol. 22, pp. 483-493.

[2] Owings ,T. M., Pavol , M. J. , Foley ,K. T. and Grabiner , M. D., "Exercise: Is it a solution to falls by elder adults?," *Journal of Applied Biomechanics*, 1999, Vol. 15, pp. 56-63.

[3] Rantz, M.J., Skubic, M., Abbott, C., Galambos, C., Pak, Y., Ho, D.K., Stone, E.E., Rui, L., Back, J., & Miller, S.J., "In-home fall risk assessment and detection sensor system,"*Journal of Gerontological Nursing,* 2013, *39*(7), 18-22.

[4] Lai, D., Taylor, S., and Begg ,R., "Prediction of foot clearance parameters as a precursor to forecasting the risk of tripping and falling," *Human Movement Science*, 2012, Vol. 31 pp. 271–283.

[5] Barrett,R., Mills,P., and Begg, R., "A systematic review of the effect of ageing and falls history on minimum foot clearance characteristics during level walking," *Gait and Posture*, 2010, Vol. 32, Issue 4, pp 429-435.

[6] Kobayashi,Y., Hobara, H., Matsushita,S., and MochimaruM., "Key joint kinematic characteristics of the gait of fallers identified by principal component analysis," *Journal of Biomechanics*, 2014, Vol. 47, Issue 10, pp. 2424.

[7] BeggR. K., Best ,R. J., Taylor,S. and Dell'OroL., " Minimum foot clearance during walking: Strategies for the minimization of trip-related falls," *Gait and Posture*. 2007.

[8] Lai, D., Begg, R., and Palaniswami, M., "Computational Intelligence in Gait Research: A Perspective on Current Applications and Future Changes," *IEEE Transactions on Information Technology in Biomedicine*, 2009, Vol. 13, No. 5, pp. 687-702.

[9] Su,F. C., Wu,W. L., Cheng,Y. M. and Chou,Y. L., "Fuzzy clustering of gait patterns of patients after ankle arthrodesis based on kinematics parameters," *Medical Engineering and Physics*, 2001, Vol. 23, pp. 83-90.

[10] Hassan,R. , Begg,R. and Taylor,S., "Fuzzy logic based recognition of gait changes due to trip related falls," *Proceedings of 27th IEEE-EMBS Conference*, 2005, pp. 4970-4973.

[11] Chiu,S. L., "Extracting Fuzzy Rules from Data for Function Approximation and Pattern Classification", Chapter 9 in *Fuzzy Information Engineering: A guided tour of applications* edited by D. Dubois, H. Prade and R. Yager, John Wiley & Sons, 1997.

[12] Begg,R., Palaniswami,M. and Owen,B., "Support vector machines for automated gait classification," *IEEE Transactions on Biomedical Engineering*, 2005, Vol. 52, No. 5, pp. 828-838.