

IMPROVING SPEECH EMOTION RECOGNITION USING ADAPTIVE GENETIC ALGORITHMS

Mohammad Hossein Sedaaghi¹, Dimitrios Ververidis², and Constantine Kotropoulos²

¹Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran

²Dept. of Informatics, Aristotle Univ. of Thessaloniki, Box 451, Thessaloniki 54124, Greece
sedaaghi@{sut.ac.ir,aiaa.csd.auth.gr},{jimver,costas}@aiaa.csd.auth.gr}

ABSTRACT

Several methods for automatic classification of utterances into emotional states have been proposed. However, the reported error rates are rather high, far behind the word error rates in speech recognition. Accordingly, there is a constant motivation for performance optimization. In this paper, self-adaptive genetic algorithms are employed to search for the worst performing features with respect to the probability of correct classification achieved by the Bayes classifier in a first stage. That is, a genetic algorithm-based implementation of backward feature selection is proposed. These features are subsequently excluded from sequential floating feature selection employing the probability of correct classification achieved by the Bayes classifier as criterion. In a second stage, self-adaptive genetic algorithms are employed to search for the worst performing utterances with respect to the same criterion. The sequential application of both stages is demonstrated to improve speech emotion recognition on the Danish Emotional Speech database.

1. INTRODUCTION

Vocal emotions constitute an important constituent of multimodal human computer interaction [1, 2]. Quantitative studies of vocal emotions have had a longer history than quantitative studies of facial expressions [3]. Several recent surveys are devoted to the analysis and synthesis of speech emotions from the point of view of pattern recognition and machine learning as well as psychology [4–7].

One approach for speech emotion analysis classifies utterances into discrete categories such as anger, happiness, sadness, surprise, neutral, etc. This is in par with neurophysiological and neuroimaging evidence suggesting that the human brain contains facial expression recognition detectors specialized for specific discrete emotions [8]. However, behavioral evidence implies that emotion categories are not entirely discrete and independent, because some emotion types tend to overlap in the sense that some types (e.g. anger and disgust) are closer than others (e.g. sadness and happiness) in emotion space. This dichotomy is evident in speech emotion classification literature, where researchers adopt either the discrete case [9–12] or work on the continuous arousal-valence space [13, 14], to mention a few.

In machine learning, let the objects be described by a vector of numerical or nominal features. If the number of features is N , there are 2^N possible feature subsets. Feature selection is a topic at the cross-section of several disciplines such as pattern recognition and machine learning, statistics, information theory, and the philosophy of science. It is essentially an optimization problem that involves searching the space of possible feature subsets to find one subset that is optimal (or near-optimal) with respect to a certain criterion [15–18]. Every feature subset selection algorithm contains two main parts: (1) the search strategy employed to select the feature subsets and (2) the evaluation method applied to test their goodness and fitness based on some criteria. Search strategies can be classified into one of the following three categories: (1) optimal, (2) heuristic, and (3) randomized. Exhaustive search is the most straightforward approach to optimal feature selection. However,

since the number of possible subsets grows exponentially, exhaustive search becomes impractical for even moderate feature numbers. The only optimal feature selection method, which avoids the exhaustive search is based on the branch and bound algorithm [19]. Sequential forward selection (SFS) and sequential backward selection (SBS) are two well-known heuristic suboptimal feature selection schemes. SFS, starting with an empty feature set, selects the best single feature and then adds it to the feature set. SBS starts with the entire feature set and at each step drops the feature whose absence decreases the performance. Combining SFS and SBS gives birth to plus l -take away r feature selection, which first enlarges the feature subset by adding l features using SFS and then deletes r features using SBS. Sequential forward floating search (SFFS) and sequential backward floating search (SBFS) are generalizations of the plus l -take away r method, where l and r are determined automatically and updated dynamically [20]. SFFS is found to dominate among 15 feature selection methods in terms of classification error and run time on a 2-class, 20-dimensional, multivariate Gaussian data set [17]. SFFS results are found comparable to those of optimal branch-and-bound algorithm, while requiring less computation time.

Feature selection can be performed with respect to properties, such as orthogonality, correlation, mutual information, etc. From the perspective of the criterion employed, feature selection methods can be distinguished as either *filters* or *wrappers*. Filters are computationally more efficient than wrapper approaches since they evaluate the goodness of selected features using criteria that can be tested quickly (e.g., reducing the correlation or the mutual information among features). This, however, could lead to non-optimal features, especially, when the features depend on the classifier. As a result, classifier performance might be poor. Wrappers train a classifier using the selected features and estimate the classification error using a validation set. Although the latter procedure is slower than filters, the selected features are usually more discriminative for the specific classifier [21, 22].

Computational studies of Darwinian evolution and natural selection have led to numerous models for computer optimization. Evolutionary algorithms have also been used for feature selection [23]. They are random search algorithms. Among them genetic algorithms (GA) comprise a subset of evolutionary algorithms focusing on the application of selection, mutation, and recombination to a population of competing problem solutions [24]. Obviously, GAs are prime candidates for random probabilistic search algorithms within the context of feature selection [25–28].

In classification, labelled examples induce a model that classifies objects into a finite set of known classes. There are three reasons for subset feature selection in conjunction with classification. First, irrelevant, non informative features may result in a classifier which is not robust. This is due to the fact that classification error does not satisfy monotonicity. Second, a large number of features implies also a large number of observations to properly design a classifier. Finally, by eliminating irrelevant features, classification time and time for data collection can be reduced. Frequently, before proceeding to speech emotion recognition subset feature selection is performed [9, 11, 29]. GAs have also been employed for feature

generation in speech emotion recognition [10].

In this paper, we employ self-adaptive GAs to further reduce the prediction error for speech emotion recognition reported in [9, 12]. Self-adaptive GAs change the probabilities of crossover and mutation during generations based on population diversity [30, 31]. They are employed to search for the worst performing features with respect to the probability of correct classification achieved by the Bayes classifier in a first stage. That is, a genetic algorithm-based implementation of BFS is proposed. These features are subsequently excluded from sequential floating feature selection employing the probability of correct classification achieved by the Bayes classifier as criterion. In a second stage, self-adaptive GAs are employed to search for the worst performing utterances with respect to the same criterion. The sequential application of both stages is demonstrated to improve speech emotion recognition on the Danish Emotional Speech database [32].

In GA literature, a binary string codes the chromosomes (i.e. features or utterances in this paper). In this binary coding, 1 implies that the feature/utterance is active and 0 implies the opposite. In this paper, another coding is employed that codes the location of active features/utterances. That is, integer values are used, which refer to the location of the worst features/utterances that should be excluded from further consideration. Definitely, the number of the worst features are much less than the best ones. Therefore, instead of having a lengthy binary stream, we have a very short integer stream that can easily be interpreted.

The outline of the paper is as follows. Section 2 briefly describes GAs. The proposed method is outlined in Section 3. Experimental results are demonstrated in Section 4 and conclusions are drawn in Section 5.

2. GENETIC ALGORITHMS

In this Section, the operators of the self-adaptive GAs are briefly described. In the following, genes refer to integer-valued elements of chromosomes (i.e. strings of genes encoding individuals). Instead of searching for best genes, we are interested in seeking the worst ones. An integer matrix \mathbf{P} of dimensions $N_p \times N_w$ is defined whose element P_{ij} codes the feature index of the j th worst gene of the i th individual (chromosome). P_{ij} admits an integer value in the range $[1, N]$, where N is the number of features in the first stage or the number of utterances in the second stage. N_p and N_w are predefined.

Let us define the notion of *population diversity* as the normalized square root of the sum of differences between any two distinct rows of the population matrix, i.e.

$$D = \frac{2}{N_p(N_p - 1)} \sum_{i=1}^{N_p-1} \sum_{j=i+1}^{N_p} \sqrt{(\mathbf{p}_i - \mathbf{p}_j)(\mathbf{p}_i - \mathbf{p}_j)^T} \quad (1)$$

where \mathbf{p}_i is a row vector that represents the i th chromosome. To avoid misunderstandings, inner products are employed in (1).

2.1 Initial population

In general, the initial population is generated randomly. To do so, a uniform random number generator fills in \mathbf{P} with integers in the desired range. P_{ij} are checked for uniqueness inside each chromosome. Typical values of N_p could be 50, 100, 200. The default value of N_p is 100. We have also made experiments with $N_p = 50, 200$ without noticing any significant difference. Let N_{iter} denote the number of iterations. N_{iter} typically admits values 50, 100, and 200. However, the larger N_{iter} is, the higher the chance to find the optimal value is, but at the expense of more computational time. If $N_{iter} < 50$, then there will be no reliable result. If self-adaptive GAs are not employed, it is more probable to get a null diversity when N_{iter} is large. The latter happens because, the dominant chromosome most probably fills in all rows of \mathbf{P} after some iterations.

2.2 Selection

The selection strategy is cross generational and differs from traditional selection. In traditional selection, the fittest genes have more chance to survive. However, in cross generational selection, additional random chromosomes are appended in \mathbf{P} . The number of new chromosomes could be N_p or a fraction of N_p . In our experiments another N_p chromosomes are randomly generated, and N_p out of the $2N_p$ worst individuals with respect to the fitness criterion are given a chance to survive in the next generations.

The evaluation procedure for the fitness of population is the repeated ψ -fold cross validation (i.e. [33]). We preserve the N_p worst chromosomes for the next operations.

2.3 Crossover

We apply a simple multi-point crossover operator [34]. The number of points and also their positions are determined randomly for any pair of candidate parents for crossover. The probability of the crossover is determined by the status of population diversity. We call it self-adaptive crossover.

2.4 Mutation

A single-point binary mutation at point k (i.e., the k th bit is toggled) is performed [34]. The probability of mutation is also determined by the population diversity. We call it self-adaptive mutation. The choice of the crossover rate is not critical compared to the probability of mutation. A large value of the probability of mutation will not allow for optimizing the fitness function and the GA will perform a random search. On the contrary, a small value will not allow the search to escape from local minima.

3. THE PROPOSED METHOD

The outline of the proposed method is as follows.

1. Generate the matrix \mathbf{P} of size $N_p \times N_w$, for $N_p = 100$. For feature trimming, N_w may vary from 1 to N_f , where N_f denotes the number of the features. In the experiments for feature trimming reported in Section 4, $N_w = 1$. For utterance trimming, $N_w \in \{1, \dots, N_u\}$, where N_u denotes the number of the utterances. In the experiments for utterance trimming reported in Section 4, $N_w = 3$.
2. Assure that there are no repetitions inside each row as well as between rows.
3. Evaluate the fitness of the initial population.
4. Repeat the following steps, until all population chromosomes have been examined (i.e. the maximum generation is reached). Also control the diversity of the population. If it reaches 0, then quit the loop.
5. Start a loop. Generate another N_p chromosomes in the selection stage and attach them to the previous population. Then, evaluate their fitness. Select the worst N_p chromosomes.
6. Calculate the diversity of the population and select probabilities of the crossover and mutation operators. If the diversity is more than a threshold, then assign a minimum value to both probabilities (e.g. 0.5 to crossover and 0.01 to mutation). Let T_{min} and T_{max} define two thresholds. If $D < T_{min}$, then increase the probabilities of crossover and mutation. If $D > T_{max}$, then decrease them. Otherwise, don't modify them. In our experiments, T_{min} and T_{max} were defined as 0.1 and 0.95, respectively.
7. Apply the crossover operation to randomly selected parents pairs.
8. Apply the mutation to randomly selected parents.
9. Repeat the loop (i.e., jump to step 4).
10. After the GA has converged, then remove the worst features/utterances from the dataset.
11. Evaluate the remaining features using the SFFS algorithm with criterion the probability of correct classification achieved by the Bayes classifier, when the features are modelled by a multivariate Gaussian probability density function. If some utterances

are excluded then SFFS is applied on the retained utterances and the probability of correct classification of the Bayes classifier is estimated by the repeated ψ -fold cross validation.

Fig. 1 illustrates the flowchart of the proposed method.

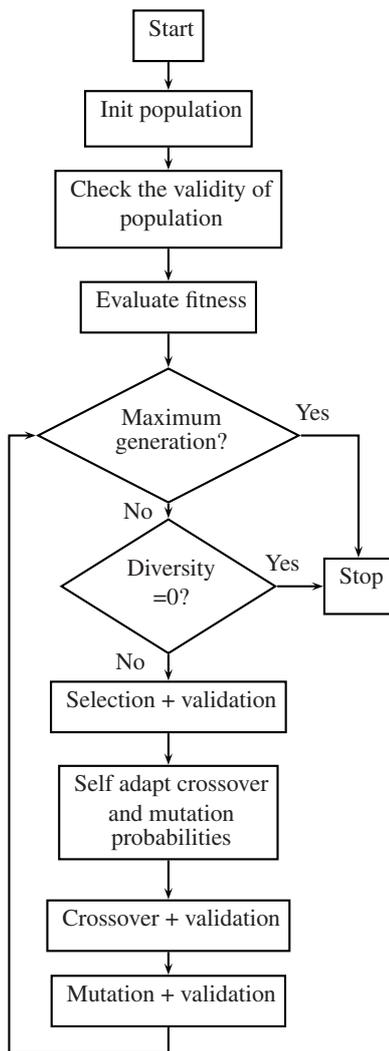


Figure 1: Flowchart of the proposed method.

4. EXPERIMENTAL RESULTS

Emotional speech data from Danish Emotion Speech (DES) [32] are employed. The recordings correspond to speech expressed by 2 male and 2 female actors under 5 emotional states such as anger, happiness, neutral, sadness, and surprise. The speech data consist of 2 words, 9 sentences, and 2 paragraphs. Overall, 1160 utterances have been used. Gender information has not been exploited. The basis for our experiments is the results reported in [9, 12].

The statistical features employed in this study are grouped in several classes as is explained in the sequel. Throughout the analysis following, the features are referenced by their corresponding indices.

1. Formants features: The set of formants features indexed by 1-16 is comprised by the statistical properties of the 4 formant frequency contours. 1. - 4. Mean value of the first, second, third, and fourth formant 5. - 8. Maximum value of the first, second, third, and fourth formant 9. - 12. Minimum value of the first, second, third, and fourth formant 13. - 16. Variance of the first, second, third, and fourth formant

2. Pitch features: The pitch features indexed by 17-51 are statistics of the pitch frequency contour. 17. - 21. Maximum, minimum, mean, median, interquartile range of pitch values. 22. Pitch existence in the utterance expressed in percentage (0-100%). 23. - 26. Maximum, mean, median, interquartile range of durations for the plateaux at minima. 27. - 29. Mean, median, interquartile range of pitch values for the plateaux at minima. 30. - 34. Maximum, mean, median, interquartile range, upper limit (90%) of durations for the plateaux at maxima. 35. - 37. Mean, median, interquartile range of the pitch values within the plateaux at maxima. 38. - 41. Maximum, mean, median, interquartile range of durations of the rising slopes of pitch contours. 42. - 44. Mean, median, interquartile range of the pitch values within the rising slopes of pitch contours. 45. - 48. Maximum, mean, median, interquartile range of durations of the falling slopes of pitch contours. 49. - 51. Mean, median, interquartile range of the pitch values within the falling slopes of pitch contours.
3. Energy (intensity) features: The energy features indexed by 52-85 are statistics of the energy contour. 52. - 56. Maximum, minimum, mean, median, interquartile range of energy values. 57. - 60. Maximum, mean, median, interquartile range of durations for the plateaux at minima. 61. - 63. Mean, median, interquartile range of energy values for the plateaux at minima. 64. - 68. Maximum, mean, median, interquartile range, upper limit (90%) of duration for the plateaux at maxima. 69. - 71. Mean, median, interquartile range of the energy values within the plateaux at maxima. 72. - 75. Maximum, mean, median, interquartile range of durations of the rising slopes of energy contours. 76. - 78. Mean, median, interquartile range of the energy values within the rising slopes of energy contours. 79. - 82. Maximum, mean, median, interquartile range of durations of the falling slopes of energy contours. 83. - 85. Mean, median, interquartile range of the energy values within the falling slopes of energy contours.
4. Spectral features: The spectral features indexed by 86-113 is the energy content of certain frequency bands divided to the length of the utterance. 86. - 93. Energy below 250, 600, 1000, 1500, 2100, 2800, 3500, 3950 Hz. 94. - 100. Energy in the frequency bands 250 - 600, 600 - 1000, 1000 - 1500, 1500 - 2100, 2100 - 2800, 2800 - 3500, 3500 - 3950 Hz. 101. - 106. Energy in the frequency bands 250 - 1000, 600 - 1500, 1000 - 2100, 1500 - 2800, 2100 - 3500, 2800 - 3950 Hz. 107. - 111. Energy in the frequency bands 250 - 1500, 600 - 2100, 1000 - 2800, 1500 - 3500, 2100 - 3950 Hz. 112. - 113. Energy ratio between the frequency bands (3950 - 2100) and (2100 - 0) and between the frequency bands (2100 - 1000) and (1000 - 0).

Then the following features are discarded: 8, 23-29, 33-34, 41, 48, 57-63, 67, 75, 82, 105. Thus, 90 features out of 113 are retained for further consideration as in [12].

The following features have been selected as the most discriminating ones by the Bayes classifier using SFFS, when 10% of the utterances were used for testing and there are 10 repetitions of 10-fold cross-validation:

1. Feature 52: Maximum of energy values.
2. Feature 21: Interquartile range of pitch values.
3. Feature 17: Maximum of pitch values.
4. Feature 39: Mean duration of the rising slopes of pitch contours.
5. Feature 20: Median of pitch values.

The following 5 features have been selected as the most discriminating ones by the Bayes classifier, when feature 2 and utterances 1132-1135 are excluded based on the results of the GA:

1. Feature 53: Minimum of energy values.
2. Feature 21: Interquartile range of pitch values.
3. Feature 113: Energy ratio between the frequency bands (2100 - 1000) and (1000 - 0).
4. Feature 39: Mean duration of the rising slopes of pitch contours.
5. Feature 1: Mean value of the first formant.

We have run classical, adaptive, and self-adaptive GAs to investigate the possibility of improving speech emotion recognition by excluding the worst performing features, before applying SFFS. Among them, the results for self-adaptive were found to be promising.

Fig. 2 illustrates how well the self-adaptive GA controls the diversity of the population along generations in one of the experiments within $N_{iter} = 50$ iterations.

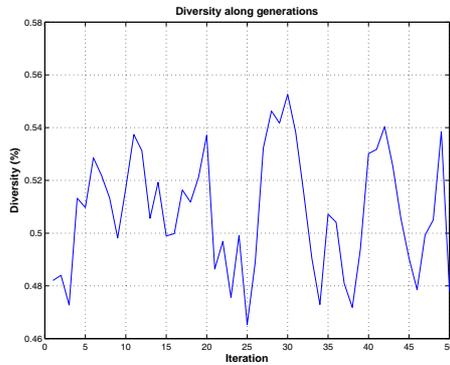


Figure 2: Linearly scaled diversity in the range [0, 1] along generations.

Table 1 presents confusion matrix from subjective human evaluation [32]. The utterances are correctly identified with an average rate of 67%. ‘‘Surprise’’ and ‘‘Happiness’’ are often confused as well as ‘‘Neutral’’ and ‘‘Sadness’’.

Table 1: Confusion matrix from subjective human evaluation [32].

Stimuli	Correctly classified responses (%)				
	Anger	Happ.	Neutral	Sadness	Surprise
Anger	75.1	4.5	10.2	1.7	8.5
Happiness	3.8	56.4	8.3	1.7	29.8
Neutral	4.8	0.1	60.8	31.7	2.6
Sadness	0.3	0.1	12.6	85.2	1.8
Surprise	1.3	28.7	10.0	1.0	59.1
Total error rate (%)	32.7				

Table 2 shows the confusion matrix for speech emotion recognition using the Bayes classifier with SFFS [9] for 10 cross-validation repetitions.

Table 2: Confusion matrix for the Bayes classifier with SFFS when cross-validation repetitions are limited to 10 [9].

Stimuli	Correctly classified responses (%)				
	Anger	Happ.	Neutral	Sadness	Surprise
Anger	37.95	20.73	11.69	11.22	18.41
Happiness	16.58	32.83	13.50	11.56	25.53
Neutral	5.97	5.67	45.25	35.62	7.49
Sadness	2.76	4.34	23.29	63.01	6.60
Surprise	14.12	19.90	7.28	12.81	45.89
Total error rate (%)	55.02				

The total error rate (i.e., the average prediction error) obtained when features are excluded with and without excluding utterances is plotted in Fig. 3. The events in Fig. 3 are decoded as follows.

- Event 1: Total error rate without applying the proposed method.
- Event 2: Total error rate with the proposed method when feature 112 is excluded.
- Event 3: Total error rate with the proposed method when feature 111 is excluded.
- Event 4: Total error rate with the proposed method when feature 104 is excluded.
- Event 5: Total error rate with the proposed method when feature 2 is excluded.
- Event 6: Total error rate with the proposed method when utterances 1132-1135 but no feature are excluded.
- Event 7: Total error rate with the proposed method when utterances 1132-1135 and feature 112 are excluded.
- Event 8: Total error rate with the proposed method when utterances 1132-1135 and feature 111 are excluded.
- Event 9: Total error rate with the proposed method when utterances 1132-1135 and feature 104 are excluded.
- Event 10: Total error rate with the proposed method when utterances 1132-1135 and feature 2 are excluded.

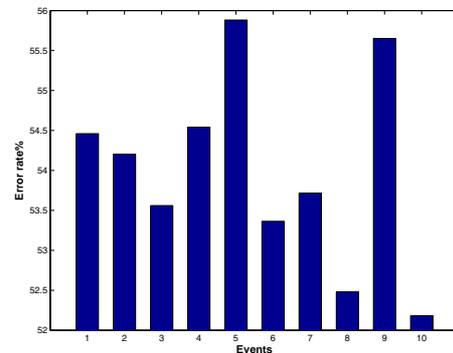


Figure 3: Comparison of total error rates.

Table 3 demonstrates the confusion matrix, when utterances 1132-1135 and feature 2 (i.e. the mean value of the second formant) have been excluded before emotional speech recognition. It is seen that the probability of correct decisions for anger, neutral, sadness, and surprise is slightly increased. Therefore, the first results reported are promising, because the proposed algorithm is able to detect the worst features and the most problematic utterances.

Table 3: Confusion matrix when the mean value of the second formant and utterances 1132-1135 are removed by the GA from subsequent classification.

Stimuli	Correctly classified responses (%)				
	Anger	Happ.	Neutral	Sadness	Surprise
Anger	44.52	18.18	10.77	15.00	11.53
Happiness	20.11	30.48	12.96	15.23	21.22
Neutral	4.52	3.63	52.19	34.24	5.42
Sadness	5.10	1.68	19.26	69.43	4.53
Surprise	15.05	16.43	7.70	18.34	42.48
Total error rate (%)	52.18				

5. CONCLUSION AND FUTURE WORK

We have applied self-adaptive GAs to increase the probability of correct classification in emotional speech recognition when the Bayes classifier with feature subset selection is used. Future work will address smoothing of extracted features, before emotional speech recognition.

Acknowledgement

This work was carried out during the tenure of a MUSCLE Internal fellowship.

REFERENCES

- [1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80, 2001.
- [2] M. Pantic and L. J. M. Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390, 2003.
- [3] A. Jaimes and N. Sebe. Multimodal human computer interaction: A survey. In *Proc. IEEE Int. Workshop Human Computer Interaction in conjunction with ICCV*. Beijing, China, Oct 2005.
- [4] P.-Y. Oudeyer. The production and recognition of emotions in speech: features and algorithms. *Int. J. Human-Computer Studies*, 59:157–183, 2003.
- [5] K. R. Scherer. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40:227–256, 2003.
- [6] P. N. Juslin and P. Laukka. Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological Bulletin*, 129(5):770–814, 2003.
- [7] D. Ververidis and C. Kotropoulos. Emotional speech recognition: Resources, features, and methods. *Speech Communication*, 48(9):1162–1181, 2006.
- [8] J. M. Susskind, G. Littlewort, M. S. Bartlett, J. Movellan, and A. K. Anderson. Human and computer recognition of facial expressions of emotion. *Neuropsychologia*, 45:152–162, 2007.
- [9] D. Ververidis and C. Kotropoulos. Automatic speech classification to five emotional states based on gender information. In *Proc. XII European Signal Processing Conf.*, volume 1, pages 341–344. Vienna, Austria, Sep 2004.
- [10] B. Schuller, D. Arsić, F. Wallhoff, M. Land, and G. Rigoll. Bioanalog acoustic emotion recognition by genetic feature generation based on low-level-descriptors. In *Proc. Int. Conf. Computer as Tool (EUROCON)*, pages 1292–1295, 2005.
- [11] B. Schuller, R. Jiménez, G. Rigoll, and M. Lang. Meta-classifiers in acoustic and linguistic feature fusion-based affect recognition. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, volume 1, pages 325–328, 2005.
- [12] D. Ververidis and C. Kotropoulos. Fast sequential floating forward selection applied to emotional speech features estimated on DES and SUSAS data collections. In *Proc. XIV European Signal Processing Conf.*, 2006.
- [13] Z. Hammal, B. Bozkurt, L. Couvreur, U. Unay, A. Caplier, and T. Dutoit. Passive versus active: Vocal classification system. In *Proc. XIII European Signal Processing Conf.* Antalya Turkey, Sep 2005.
- [14] J. Kim, E. André, M. Rehm, T. Vogt, and J. Wagner. Integrating information from speech and physiological signals to achieve emotional sensitivity. In *Proc. 9th European Conf. Speech Communication and Technology*, 2005.
- [15] P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall, 1993.
- [16] F. J. Ferri, P. Pudil, M. Hatef, and J. Kittler. Comparative study of techniques for large scale feature selection. In J. E. Moody, S. J. Hanson, and R. L. Lippmann, editors, *Pattern Recognition in Practice IV*, pages 403–413, 1994.
- [17] A. K. Jain and D. Zongker. Feature selection: evaluation, application, and small sample performance. *IEEE Trans. Pattern Anal., Machine Intell.*, 19(2):153–158, 1997.
- [18] M. Dash and H. Liu. Feature selection for classification. *Intelligent Data Analysis*, 1(3):131–156, 1997.
- [19] P. Somol, P. Pudil, and J. Kittler. Fast branch & bound algorithms for optimal feature selection. *IEEE Trans. Pattern Anal., Machine Intell.*, 26(7):900–912, July 2004.
- [20] P. Pudil, J. Novovicova, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, 1994.
- [21] G. John, R. Kohavi, and K. Phleger. Irrelevant features and the feature subset problem. In W. W. Cohen and H. Hirsh, editors, *Proc. 11th Int. Conf. Machine Learning*, pages 121–129. San Francisco, CA, Morgan Kaufmann, 1994.
- [22] R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324, 1997.
- [23] W. Siedlecki and J. Sklansky. A note on genetic algorithm for large-scale feature selection. *Pattern Recognition Letters*, 10:335–347, 1989.
- [24] D. Goldberg, editor. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley, Reading, MA, 1989.
- [25] M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn, and A. K. Jain. Dimensionality reduction using genetic algorithms. *IEEE Trans. Evolutionary Computation*, 4(2):164–171, 2000.
- [26] Z. Sun, G. Bebis, and R. Miller. Object detection using feature subset selection. *Pattern Recognition*, 37:2165–2176, 2004.
- [27] D. P. Muni, N. R. Pal, and J. Das. Genetic programming for simultaneous feature selection and classifier design. *IEEE Trans. Systems, Man & Cybernetics-Part B: Cybernetics*, 36(1):106–117, February 2006.
- [28] E. Zio, P. Baraldi, and N. Pedroni. Selecting features for nuclear transients classification by means of genetic algorithms. *IEEE Trans. Nuclear Science*, 53(3):1479–1493, June 2006.
- [29] J. Wagner, J. Kim, and E. André. From physiological signals to emotion: Implementing and comparing selected methods for feature extraction and classification. In *Proc. 2005 IEEE Int. Conf. Multimedia and Expo*, 2005.
- [30] B. A. Julstrom. Adaptive operator probabilities in a genetic algorithm that applies three operators. In *Proc. 1997 ACM Symp. Applied Computing*, pages 233–238. San Jose, CA, 1997.
- [31] P. J. Angeline. Adaptive and self-adaptive evolutionary computations. In M. Palaniswami and Y. Attkiuzel, editors, *Computational Intelligence: A Dynamic Systems Perspective*, pages 152–163. IEEE Press, 1995.
- [32] I. S. Engberg and A. V. Hansen. Documentation of the danish emotional speech database (des), 1996.
- [33] P. Burman. A comparative study of ordinary cross-validation, ψ -fold cross-validation and the repeated learning-testing methods. *Biometrika*, 76(3):503–514, 1989.
- [34] J. Joines. The genetic algorithm optimization toolbox for MATLAB.