

Technical Report

TO: University of Zabol

FROM: Najme Ghanbari

Research Project Number: 92 -16

DATE: November 25, 2015

SUBJECT: The GSA Algorithm at Two Methods of Feature Selection and Weighting Features to Improve the Recognition Rate of Persian Handwrite Digits with Fuzzy Classifier

Introduction: One of the most efficient ways to improve accuracy in a handwritten digits recognition system is selecting optimal features among the entire set of extracted features. In this project, the two methods proposed to increase the recognition rate of Persian handwritten digits. In two methods Gravitational search algorithm (GSA) was used. Binary Gravitational search algorithm (BGSA) proposed for selecting optimal features. On the other hand, one of the aims of this project is to evaluate the increased recognition rate of Persian handwritten digits by using the BGSA. Also, in different proposed way, Persian handwritten recognition rate improved. In this method instead of choosing some of the features by BGSA, one random weight has been assigned to each feature in order to improve recognition rate. Finding the Weight vector by mathematical and statistical computational methods is very difficult. This Weight vector is obtained with real version of GSA (RGSA).

Methods: Through using the binary Gravitational search algorithm (BGSA) version in the proposed method, we select the optimal features and find the best features set among the whole extracted features. Not only the number of features and computational burden reduces, but also recognition rate will significantly increase. Fuzzy classifier is also used for the classification in this project. Fitness function in BGSA is the number of fuzzy classifier errors which aims to minimize this value. Also in another proposed method, instead of choosing some of the features by BGSA, one random weight has been assigned to each feature in order to improve recognition rate. This Weight vector is obtained with RGSA. Mr. Khosravi's database has been used in this project, because it is the better database than the other ones in terms of data volume and data diversity.

Results: Both methods significantly improve the recognition rate. There is always a tradeoff between training and testing stages. That is, if the high speed decision produces and low computational costs achieves in practical stage, the costs spent to appropriately select the patterns and features will be much more. The more additional costs for training, the compensated costs for practice stage would be more. The obtained results of using BGSA and RGSA confirmed that these algorithms enjoy proper performance. Furthermore, the results of recognition are without any preprocessing and post processing operations; so they obtained recognition rate is appropriate. These methods can also be implemented with other algorithms based on population.