UNIVERSITI TEKNOLOGI MARA

A GUIDED HYBRID K-MEANS AND GENETIC ALGORITHM MODELS FOR CHILDREN HANDWRITING LEGIBILITY PERFORMANCE ASSESSMENT

NORZEHAN BINTI SAKAMAT

Thesis submitted in fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information Technology and Quantitative Sciences)

Faculty of Computer and Mathematical Sciences

February 2021

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the results of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any degree or qualification.

I, hereby, acknowledge that I have been supplied with the Academic Rules and Regulations for Post Graduate, Universiti Teknologi MARA, regulating the conduct of my study and research.

Name of Student   :   Norzehan Binti Sakamat

Student I.D. No.   :   2011477874

Programme   :   Doctor of Philosophy (Information Technology and Quantitative Sciences) - CS990

Faculty   :   Faculty of Computer and Mathematical Sciences

Thesis Title   :   A Guided Hybrid K-Means and Genetic Algorithm Models for Children Handwriting Legibility Performance Assessment

Signature of Student   :   ....

Date   :   February 2021

ABSTRACT

Assessing and predicting children handwriting legibility performance is necessary for providing early interventions to those with handwriting difficulties. Thus, producing a good and reliable computerized handwriting assessment instruments will depend heavily on selection of appropriate handwriting features, handwritten recognition methods and clustering methods. Offline handwritten recognition method is challenging due to the fact that individual handwriting produces variations of handwriting shape, style and orientation and the input are in static form. This research proposes to develop and analyse an offline handwriting recognition instrument performance. The instruments combine's observations, feature extraction methods and clustering methods which are expected to produce predictive results of high agreement with human experts based on evaluation of selected individually handwritten alphabets. Four handwriting components have been identified which are time completion, readability, size consistency and shape formation. Time completion was calculated by observing the number of alphabets completed within the specified time. Readability was detected using a free online optical character recognition application called Aeosoft. Size was extracted using Extreme Point Detection algorithm and Hit or Miss Transformation method was used to extract the stroke formation pattern. K-Means algorithm a popular efficient clustering techniques and genetic algorithm a widely used evolutionary algorithm and known for its adaptive nature were combined to determine the level of handwriting legibility for each child. The hybridization of the two methods were proposed due to K-Means weaknesses which are predicted that it will not produce the expected results for this study. Order of input data and rescaling the input data for standardization influence K-Means in giving accurate results. The iterative nature of K-Means and random initialization of centroids which leads the algorithm to stick in a local optimum and unable to converge to optimum results, are another weakness. The combined method is called Hybrid K-MeansCGA. Modifications of K-Means structures were done by inserting genetic algorithm operators and tuning the population. This study will also tune the generation size to see whether it have an impact on producing high agreement result with human experts. The utilization of populations has been a commonly used strategy in tuning GA when it did not perform well, however, the studies of tuning generations size in GAs to find the best solution were rarely done. Euclidean Distance, Pearson Correlation and Matching Matrix were used to measure the performance of the feature extraction and clustering methods. Recognition software achieved 87.14%, EPD algorithm achieved 73.57% and HMT algorithm achieved 74.30% prediction accuracy with OTs. While hybrid K-MeansCGA combination of fix population size=100 and various size of generation performs better than general K-Means algorithm and hybrid K-MeansCGA combination of fix generation size=100 and various size of populations. Hybrid K-MeansCGA with generation=150 and population=100 results in prediction accuracy scores of 87% with teachers and 85% prediction accuracy with OTs. Findings shows that by implementing different sizes of generations can improve the clustering results, thus verifies the statement given by the natural evolution theories that generations of species do have great impact in producing the most fit individuals. This research has achieves its objective as the combined methods are reliable instruments that best imitate the assessment decisions of occupational therapist who are the qualified professionals in treating issues related to the development of handwriting among children.

ACKNOWLEDGEMENT

In the name of Allah, The Most Gracious, Most Merciful.

Alhamdulillah with His blessing I am able to complete this long and challenging journey. First and foremost, I would like to express my sincere gratitude to my advisor Dr Noor Elaiza Abd Khalid for the continuous support of my studies, for her constant support, inspiration and immense knowledge. Her support and guidance have helped me go through all obstacles in finishing this research. I would also like to extend my sincere gratitude to Dr Marina Yusoff, my second supervisor, for her kind understanding spirit and motivational support. My sincere appreciation to Dr Sharifah Aliman, the Head of Postgraduate Studies FSKM, for her relentless encouragement that keeps me going till the end of the journey. Not forgetting, Inda Ishadah Nazrul Azhar, who is part of my research team, for the stimulating discussions and technical assistance. May Allah bless all of you always.

Deepest appreciation to the Principal of Brainy Bunch Elementary International Islamic School, Encik Muhammad Ali Nordin, all the teachers and lower elementary students, thank you for willingly participate in this research. To the staff of Centre of Occupational Therapy, Faculty of Health Sciences UiTM: Senior Lecturer, Mr Chandra Kannan K. Thanapalan; Clinical Instructor, Mr Muhammad Radhi Rahimi Abu Bakar and Mrs Nurul Izzah Wahidul Azam, thank you for all the guidance and assistance in making this research a success.

I am greatly indebted to my family: my mother, Aminah Md Saff ; husband, Johari Jihat; to my childrens: Muhammad Hariz Iman, Nur Alyaa Qaysara, Nur Aisha Sofea, Nur Aisha Umaira and Muhammad Thariq; thank you for all your love, unwavering support and sacrifices. Without your tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study. Sincere gratitude to Ustaz Sharin Abdul Rahman for all the spiritual guidance and advice when the road I am trudging seems hard and impossible to get through. To my dearest best friends Afiza and Zilah, Izan, Nana, Anom, Fairuz, Azhar, Fazila and Izyan, thank you for being by my side from the start of the journey till the end, thank you for being true friends and a friend in need, I am forever grateful for it. Not forgetting my UUM friends: Ima, Rizal, Lela, Fidz and Yati, thank you for all the joy and laughter; you open the other side of me that i have long forgotten. Last but not least, my appreciation goes to my siblings, uncle, aunties and cousins for all the moral support in this long journey. May Allah reward all of you with goodness in this world and the hereafter.

TABLE OF CONTENT