RGC Ref. No.:
UGC/FDS16/E10/19
(please insert ref. above)

RESEARCH GRANTS COUNCIL
COMPETITIVE RESEARCH FUNDING SCHEMES FOR THE LOCAL SELF-FINANCING DEGREE SECTOR

FACULTY DEVELOPMENT SCHEME (FDS)

Completion Report
(for completed projects only)

Submission Deadlines:
1. *Auditor's report with unspent balance, if any: within **six** months of the approved project completion date.*
2. *Completion report: within **12** months of the approved project completion date.*

Part A: The Project and Investigator(s)

1. Project Title

On effective utilization of small training sample for computer short answer grading based on semi-supervised clustering algorithms

2. Investigator(s) and Academic Department(s) / Unit(s) Involved

Research Team	Name / Post	Unit / Department / Institution
Principal Investigator	LUI Kwok Fai (Professor)	School of Science and Technology, Hong Kong Metropolitan University
Co-Investigator(s)	NG Sin Chun (Resigned)	School of Science and Technology, Hong Kong Metropolitan University
Others		

3. Project Duration

	Original	Revised	Date of RGC / Institution Approval *(must be quoted)*
Project Start Date	1 Jan 2020	1 Jan 2020	11 Aug 2021
Project Completion Date	31 Dec 2021	30 Jun 2022	
Duration *(in month)*	24 months	30 months	
Deadline for Submission of Completion Report	31 Dec 2022	30 Jun 2023	

4.4 Please attach photo(s) of acknowledgement of RGC-funded facilities / equipment.

N/A

Part B: The Final Report

5. Project Objectives

5.1 Objectives as per original application

1. To design and implement a framework of computer short answer grading with semi-supervised clustering based on evolutionary algorithms.
2. To model graded training sample on ordinal grading scale as clustering constraints and to formulate objective functions accordingly for multi-objective optimization.
3. To develop an effective mechanism for maximizing the utilization of manual grading effort.
4. To develop a prototype computer short answer grading system based on the integration of the outcomes of the above objectives #1 to #3.

5.2 Revised objectives

Date of approval from the RGC: N/A

Reasons for the change: N/A

5.3 Realisation of the objectives
(Maximum 1 page; please state how and to what extent the project objectives have been achieved; give reasons for under-achievements and outline attempts to overcome problems, if any)

The grading examples for training automated short answer grading models are sampled from submitted answers and annotated by human graders. This preparation task, in contrast to most other short assessment types, often suffers from effort wastage in producing redundant examples. To improve the appeal of automated short answer grading, optimizing the preparation task is desirable. Semi-supervised learning exploits any information latent in a large amount of unannotated data to increase the utility of a small sample of annotated data. The understanding of applying semi-supervised learning in training short answer grading models was considered inadequate.

Objective #1

A framework for automated short answer grading based on multi-objective evolutionary

clustering algorithm was designed, built, and evaluated. The clustering algorithm explores structures in the unannotated answers to identify representative examples. The evolutionary computing approach supports iterative querying for annotation and updating of the grading model. The cluster analysis and annotation are divided into phases of convergence. The motivation is to exploit the information in the annotated answers to identify other optimized examples such as borderline and anomalous answers. The major components of the framework, namely, compact semantic representations of short answers, cluster quality metrics, evolutionary operators, and the various essential mechanisms of evolutionary computing their effects on sampling useful examples were studied. In particular, several genotype encoding schemes based on cluster membership, cluster hierarchy, and minimum spanning trees were implemented and compared.

Objective #2

In this part of the project, the grades were considered to be quantitatively comparable. First, the framework was enhanced for an ordinal grading scale. Second, the definitions of representative and borderline answers were enriched according to the new metric of comparing different grades. In particular, the metrics used to weight the degree of uncertainty of borderline answers was formulated and tested. Another performance evaluation metric was considered to differentiate different cases of grading mistakes and a more accurate reflection of the performance.

Objective #3

The example preparation task was thoroughly analyzed for improving the utility of every manual annotation. The density peak clustering algorithm was adopted and evaluated for the grading model for this part of the project. The motivation was to resolve the heavy computation cost of evolutionary clustering and to further enrich the sampling of other optimized examples. In addition, another semi-supervised learning approach known as active learning was evaluated as an alternative. This approach was used to investigate more effective ways to exploit the information in the annotated answers, and in particular the borderline examples. An extensive evaluation based on three real-life gold standard datasets was carried out.

Objective #4

The practical use of the semi-supervised short answer grading approach was evaluated. An interactive short answer grading system was designed and implemented as a web-based platform. The core of the system was a grading model that represented the understanding of the field resulting from the previous objectives.

5.4 Summary of objectives addressed to date

Objectives *(as per 5.1/5.2 above)*	**Addressed** *(please tick)*	**Percentage Achieved** *(please estimate)*
1. To design and implement a framework of computer short answer grading with semi-supervised clustering based on evolutionary algorithms.	✓	100%
2. To model graded training sample on ordinal grading scale as clustering constraints and to formulate objective functions accordingly	✓	100%

for multi-objective optimization.		
3. To develop an effective mechanism for maximizing the utilization of manual grading effort.	✓	100%
4. To develop a prototype computer short answer grading system based on the integration of the outcomes of the above objectives #1 to #3.	✓	100%

6. Research Outcome

6.1 Major findings and research outcome
(Maximum 1 page; please make reference to Part C where necessary)

This project has made significant contribution to the understanding of useful and redundant grading examples, to the distinction of their types and properties, and to the development of techniques for the optimized grading sample requirement. At writing, this is the most comprehensive study of the active learning approach of automated short answer grading. The major findings are summarized as follows.

- A framework for building automated short answer grading models based on evolutionary clustering algorithms called NDSAGF was successfully developed and evaluated. The iterative manner of first sampling representative answers for annotations and second exploiting the annotations in refining the models was tested and found significantly improved the final cluster quality. It suggested a scheme of dividing the grading budget into more batches in order to exploit human input more in the process.

  Several genotype representations were compared, and the minimum spanning trees (MST) enabled the most efficient convergence and demonstrated the effectiveness of a pre-computed structure encoded with neighbourhood information. A few formulations for exploiting the violations found in the annotations were also investigated but the differences were marginal. It suggested other potential formulations such as those based on feature subspace or query-by-committee.

  Refer to Journal Paper #1 and Conference Paper #1 for the details.

- The enhanced framework based on the density peak clustering algorithm and the active learning approach was also successfully developed and studied. The iterative sampling of potential grading examples with the enhanced framework met the expectation of improved utility of manual preparation effort and reduced computational demand. The three types of grading examples, namely, the representative reference examples, the borderline examples, and anomalous examples, and the diversity driven batch sampling, were all found to have significant roles. The amount of redundancy was lower than the baseline semi-supervised learning methods and the random sampling method in the evaluation. In the specific study on borderline examples, there were some indications that the decision boundaries were varied between a hyper-shell, which corresponds to one-class classification, and a hyper-surface, which corresponds to conventional classifications, in the grading models.

  Refer to Journal Paper #2 and Conference Paper #2 for the details.

- A short answer grading system coupled with interactive modelling and data visualization was developed and tested. The system facilitated more informed decision making on following or turning down the recommended grading examples. The test revealed that the human input through the system could further increase the performance of the active learning framework. Refer to Conference Paper #3 for the details.

6.2 Potential for further development of the research and the proposed course of action *(Maximum half a page)*

The project has found out some significant understanding of the semi-supervised learning approach for building short answer grading models. The following issues revealed in the project deserve further investigations.

- The batch selection strategy is a part of the sampling of potential grading examples in active learning. It considers the rationale for selecting representative answers, borderline answers, or other types of answers. In the literature it is referred to the balance between exploration and exploitation. The research project discovered that the more open-ended short answer questions were associated with more borderline answers and more exploitation. Understanding the relation between the suitable batch selection strategy and the characteristic of short answer questions is important.
- The decision boundaries and the borderline examples are mutually dependent. The one-class spherical neighbourhood concept assumes all features are equally important. Other shapes may be more appropriate to cater for ignoring irrelevant features. In addition, operators such as merge or split may be useful to combine or divide one-class decision boundaries when informed by additional annotations in the active learning process.
- Human input can contribute to the preparation of grading example and also the tuning of the grading model. The interactive modelling concept supported by the short answer grading system can exploit more from the presence of human. Investigation is needed to weight the benefits against the additional effort needed.

7. Layman's Summary

(Describe in layman's language the nature, significance and value of the research project, in no more than 200 words)

The global prevalence of online learning has driven a strong demand for around-the-clock automated assessment services. The technology of automated grading can efficiently process many answers according to human-prepared grading examples. The possible answers to a short answer question can vary significantly. The wide range of possible answers require a proportional number of grading examples. Manually crafting the examples is effort intensive. A more common method to prepare grading examples is to obtain samples from the answer set and present these samples to a human grader for the grades. As the examples can be unimportant or a similar one already sampled, optimizing the preparation task is critical to the practicality of automated short answer grading. This project was initiated to increase the understanding of the example preparation task, the types of examples significant to automated short answer graders, and effective methods for their identification. The major contributions of the project include (1) an iterative sampling method has been developed and evaluated, that exploiting existing examples can improve the identification of more relevant examples and reduce redundancy, (2) a thorough analysis of the definitions and the role of borderline examples, and (3) a prototype grading system has been designed and tested.

Part C: Research Output

8. Peer-Reviewed Journal Publication(s) Arising Directly From This Research Project
(Please attach a copy of the publication and/or the letter of acceptance if not yet submitted in the previous progress report(s). All listed publications must acknowledge RGC's funding support by quoting the specific grant reference.)

The Latest Status of Publications				**Author(s)** *(denote the correspond-ing author with an asterisk*)*	**Title and Journal / Book** *(with the volume, pages and other necessary publishing details specified)*	**Submitted to RGC** *(indicate the year ending of the relevant progress report)*	**Attached to this Report** *(Yes or No)*	**Acknowledged the Support of RGC** *(Yes or No)*	**Accessible from the Institutional Repository** *(Yes or No)*
Year of Publication	**Year of Acceptance** *(For paper accepted but not yet published)*	**Under Review**	**Under Preparation** *(optional)*						
2022				Andrew Kwok-Fai LUI*, Sin-Chun NG, Stella Wing-Nga, CHEUNG	Paper #1: A Framework for Effectively Utilizing Human Grading Input in Automated Short Answer Grading *International Journal of Mobile Learning and Organisation, 16*(3), pp. 266-286	Yes (2020)	Yes [Attachment #1]	Yes	Yes
		✓		Andrew Kwok-Fai LUI*, Sin-Chun NG, Stella Wing-Nga, CHEUNG	Paper #2: Improving the Appeal of Automated Short Answer Grading with Active Learning Models *Interactive Learning Environments*	No	Yes [Attachment #2]	Yes	Yes

9. Recognized International Conference(s) In Which Paper(s) Related To This Research Project Was / Were Delivered
(Please attach a copy of each conference abstract)

Month / Year / Place	Title	Conference Name	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this Report *(Yes or No)*	Acknowledged the Support of RGC *(Yes or No)*	Accessible from the Institutional Repository *(Yes or No)*
Jul-2020 Hong Kong (Virtual)	Paper #1: Effective Use of Small Amount of Graded Answers for Automated Grading of Short Answers	International Conference in Open and Innovative Education 2020	Yes (2020)	No	Yes	Yes
Dec-2020 Canberra, Australia (Virtual)	Paper #2: Entropy-based Recognition of Anomalous Answers for Efficient Grading of Short Answers with an Evolutionary Clustering Algorithm	IEEE Symposium Series on Computational Intelligence 2020 (IEEE SSCI 2020)	Yes (2020)	No	Yes	Yes
May-2022 Xian, China (Virtual)	Paper #3: An Interactive Short Answer Grading System Based on Active Learning Models	The 4th International Conference on Computer Science and Technologies in Education	No	Yes [Attachment #3]	Yes	Yes

10. Whether Research Experience And New Knowledge Has Been Transferred / Has Contributed To Teaching And Learning

(Please elaborate)

Nil

11. Student(s) Trained

(Please attach a copy of the title page of the thesis)

Name	Degree Registered for	Date of Registration	Date of Thesis Submission / Graduation
N/A			

12. Other Impact

(e.g. award of patents or prizes, collaboration with other research institutions, technology transfer, teaching enhancement, etc.)

Nil

13. Statistics on Research Outputs

	Peer-reviewed Journal Publications	Conference Papers	Scholarly Books, Monographs and Chapters	Patents Awarded	Other Research Outputs (please specify)	
No. of outputs arising directly from this research project	1 (and 1 under review)	3	0	0	Type	No.
					Nil	

14. Public Access Of Completion Report

(Please specify the information, if any, that cannot be provided for public access and give the reasons.)

Information that Cannot Be Provided for Public Access	Reasons
Nil	