AUTOMATED ANSWER PAPER EVALUATION USING DEEP LEARNING & NLP

A Project Preliminary Report

submitted by

Name	University Register No.
Pranav T N	KNR16CS046
Rahul Mohanan A K	KNR16CS047
Sourabh Subhod	KNR16CS053
Vishal V	KNR16CS059

to

THE APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE OF

Bachelor of Technology

IN

COMPUTER SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GOVERNMENT COLLEGE OF ENGINEERING KANNUR

Kannur District, Kerala State

November 2019

Institute Vision

"A globally renowned institution of excellence in engineering, education, research and consultancy."

Institute Mission

"To contribute to the society by providing quality education and training, leading to innovation, entrepreneurship and sustainable growth."

Department Vision

"To be a centre of excellence in the field of Computer Science & Engineering education and research, which extends its appreciated services to the industry and the society."

Department Mission

"To develop engineers with excellent analytic, design and implementation skills, who can expertise themselves as computer professionals, research engineers, entrepreneurs or as managers, while fulfilling their ethical and social responsibilities, in a globally competitive environment."

Department. of Computer Science and Engineering
GOVERNMENT COLLEGE OF ENGINEERING KANNUR
Kannur District, Kerala State

CERTIFICATE

November 28, 2019

*Certified that this is a bonafide record of the Project Preliminary work done by the students whose names are given below, with the title "*AUTOMATED ANSWER PAPER EVALUATION USING DEEP LEARNING & NLP*" towards the partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering of the Department of Computer Science and Engineering under A. P. J. Abdul Kalam Technological University during the year 2019-2020.*

Name	University Register No.
Pranav T N	KNR16CS046
Rahul Mohanan A K	KNR16CS047
Sourabh Subhod	KNR16CS053
Vishal V	KNR16CS059

Prof. Sajith B	**Prof. Baburaj K V**	**Dr. Rafeeque P C**
Project Guide	Project Co-ordinator	HoD, Dept. of CSE

Acknowledgements

We are extremely grateful to **Dr. V O Rejini**, the Principal, Govt. College of Engineering, Kannur for providing the necessary facilities during the course of our project work. We are deeply indebted to **Dr. Rafeeque P C**, Head of Department of CSE for his valuable suggestions and guidance. We express our sincere thanks to the project coordinator **Prof. Baburaj K V** for his support and efforts and our guide **Prof. Sajith B** for the support and constant encouragement that has been provided along with his valuable guidance for the project. We thank all the teaching and non-teaching staff, our classmates and friends for sharing their knowledge and valuable suggestions.

Abstract

Manual answer paper evaluation is time consuming. This is evident from the current KTU evaluation system. To solve this issue, we propose a system composed of a deep learning and an NLP model. The deep learning model (LSTM) performs handwriting recognition to convert the answer paper into text. The resultant text is mapped with a answer key. This new mapping is fed into an NLP model where semantic evaluation is performed and the result is generated and sent to individual students. Deep learning is used for handwriting recognition because the training data set is huge and it also provides better accuracy when RNN models are used.

Contents

List of Figures

Chapter 1

Introduction

Automation is preferred in most areas nowadays. It saves both time and cost. In the current education system, answer scripts are evaluated by hand and this is time consuming. The result declaration is in turn delayed. Handwriting recognition is needed along with semantic analysis of text for the automation process. The technologies to be used for this are Deep Learning and Natural Language Processing (NLP).

Handwriting recognition of different languages are challenging issues and receive extensive attention from researchers. In recent years, numerous handwritten datasets have been published in the field to promote the advancement of the community. In general, handwritten datasets can be divided into two categories, i.e., online and offline datasets. For example, there are offline handwritten datasets such as French paragraph dataset Rimes [1], English text dataset IAM [2], Arabic datasets of IFN/ENIT [3] and KHATT [4], Chinese dataset CASIA-HWDB [5] and HIT-MW [6]. For online handwritten datasets, there are Japanese text datasets Kondate [7] and character dataset TUAT Nakayosi_t and Kuchibue_d [8], English text dataset IAM-OnDB [9], Chinese datasets SCUT-COUCH2009 [10], CASIAOLHWDB [5], and ICDAR2013 competition set [11]. In this project, we use the IAM dataset for training our LSTM model.

Generally in an answer paper evaluation, a human grader assesses and assigns a score to a submission which is written concerning an answer's prompt. This is a laborious and tiring task for the graders. Also, human graders can be imperfect; they are susceptible to biases and irregularities based on other chores and activities they do in life [12]. Different human graders also have different grading styles and can also tend to give an overall higher grade just based on one good impression regarding a particular aspect. A computer system can overcome all these human shortcomings by uniform assessment throughout. Understanding human language is considered a laborious task due to its complexity. There are numerous ways to arrange words in a sentence. Also, words can have multiple meanings in different contexts. Therefore context-based knowledge is necessary to decipher the sentences correctly.

Our proposed system makes use of an RNN architecture model like LSTM to recognize handwritten answer sheets and convert them to text. This text is semantically analyzed using an NLP model with an answer key to generate results.

Chapter 2

Literature Review

Handwriting recognition is needed to digitize answer papers. Deep learning is preferred for handwriting recognition for better accuracy. Zhu *et al.* [13] proposed a methodology for offline text recognition in answer papers. They noticed that most existing studies and public datasets for handwritten Chinese text recognition are based on the regular documents with clean and blank background, lacking research reports for handwritten text recognition on challenging areas such as educational documents and financial bills. They focused on examination paper text recognition and construct a challenging dataset named examination paper text (SCUT-EPT) dataset, which contains 50 000 text line images (40,000 for training and 10,000 for testing) selected from the examination papers of 2,986 volunteers. The proposed SCUT-EPT dataset presents numerous novel challenges, including character erasure, text line supplement, character/phrase switching, noised background, nonuniform word size, and unbalanced text length. In their experiments, the current advanced text recognition methods, such as convolutional recurrent neural network (CRNN) exhibits poor performance on the proposed SCUT-EPT dataset, proving the challenge and significance of the dataset. Nevertheless, through visualizing and error analysis, they observed that humans can avoid vast majority of the error predictions, which reveal the limitations and drawbacks of the current methods for handwritten Chinese text recognition (HCTR). Finally, three popular sequence transcription methods, connectionist temporal classification (CTC), attention mechanism, and cascaded attention-CTC were investigated for HCTR problem. Although the attention mechanism has been proved to be very effective in English scene text recognition, its performance is far inferior to the CTC method in the case of HCTR with large-scale character set.

Manual grading of essays by humans is time-consuming and likely to be susceptible to inconsistencies and inaccuracies. In recent years, an abundance of research has been done to automate essay evaluation processes, yet little has been done to take into consideration the syntax, semantic coherence and sentiments of the essay's text together. Janda *et al.* [14] proposed a methodology for automated essay evaluation. The proposed system incorporates not just the rule-based grammar and surface level coherence check but also includes the semantic similarity of the sentences. They proposed to use Graph-based relationships within the essay's content and polarity of opinion expressions. Semantic similarity is determined between each statement of the essay to form these Graph-based spatial relationships and novel features are obtained from it. Their algorithm uses 23 salient features with high predictive power, which is less than the current systems while considering every aspect to cover the dimensions that a human grader focuses on. Fewer features helps get rid of the redundancies of the data so that the predictions are based on more representative features and are robust to noisy data. The resulting agreement between human grader's score and the system's prediction is measured using Quadratic Weighted Kappa (QWK). Their system produces a QWK of 0.793.

Chapter 3

Project Objectives

Answer script evaluation has always been a time-consuming task in our education system. It particularly gets messy for the final year students in colleges. The evaluation should be completed as soon as possible as the degree certificates are needed for most job applications. The objectives of this project are given below:

1. A handwriting recognition system is needed to convert answer papers into digital text. Answer papers need to be converted to text so that it can be used for semantic analysis with a NLP model. This technique involves the usage of a RNN architecture like a LSTM model.

2. An NLP model is needed to perform semantic analysis with a predefined key. This model performs stemming on the answer text.

The whole system should allow teachers to upload answer sheets and automatically evaluate answer sheets and generate results and send them to individual students. All this must be done in a time less than that of manual evaluation method.

Chapter 4

Relevance & Impact on Society

This project if successful will lead to a lot of other applications. It can be extended to recognize any handwritten documents and evaluate them. It could also lead to paperless examinations, thereby abiding with Green Protocol. Result declaration of students would be less time consuming and it will be a boon for final year students as they can get their pass out certificates in less time.

Chapter 5

Work Plan

5.1 Gantt Chart

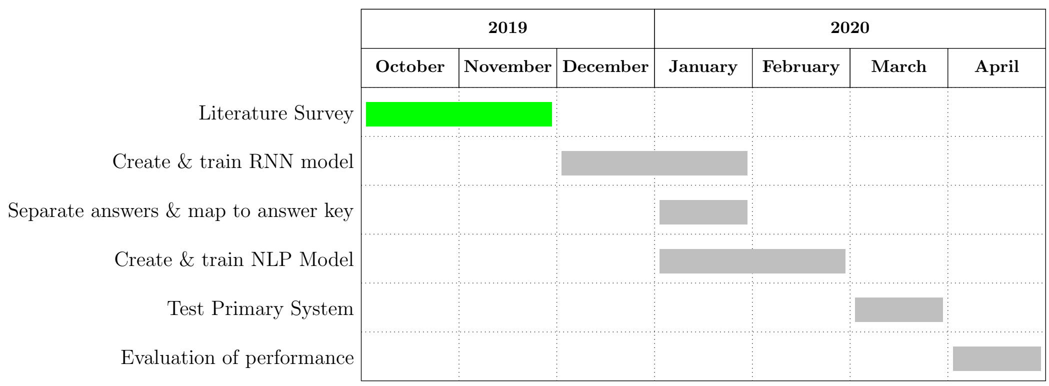

Figure 5.1: Gantt Chart

The duration of our project is from October 2019 to April 2019. We categorized our project into different phases as follows. There is a 2 month period for literature survey from October to November 2019. From December 2019 to January 2020, the RNN model will be created and trained. Parallelly, in January 2020 the separation of answers and mapping them to the answer key is done. From January 2020 to February 2020, the NLP model is created. In March 2020, the primary system is tested. And finally in April 2020, performance evaluation is done.

5.2 PERT Chart

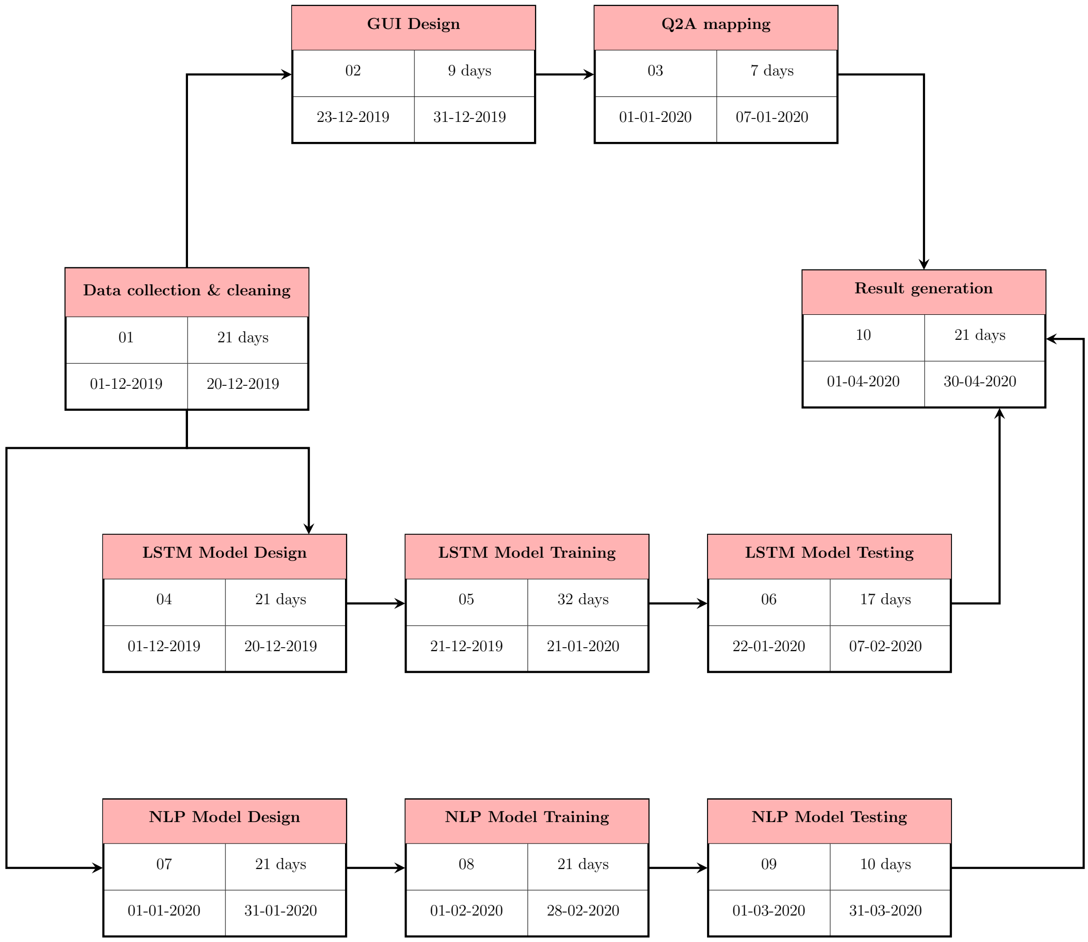

Figure 5.2: PERT chart

Initially we have a planning phase which includes data collection and cleaning phases. It is a 21 days duration phase starting from December 01 2019 to December 30 2019. At the same time, the LSTM model design will take place from Decemebr 12 2019 to December 30 2019 for a total of 21 days. The GUI design is given 9 days from December 23 2019 to December 31 2019. The LSTM model training takes place during December 21 2019 to January 21 2020. Parallelly, the Question to Answer (Q2A) key mapping also takes place during January 01 2020 to January 07 2020 for a total of 7 days. The NLP model design, training and testing takes place during the months of January 2020 to March 2020. Finally, the result generation is done during April 2020 for a total of 21 days.

Chapter 6

Functional & Non-Functional Requirements

6.1 Functional Requirements

- The system must provide the teachers with a GUI to upload answer paper and the answer key.
- The system must convert the handwritten text in answer scripts to digital text.
- The system must separate answers from the recognized text and map them to each question.
- The system must perform answer paper evaluation based on the digital text extracted and the answer key.

6.2 Non-Functional Requirements

- The system shall be able to perform evaluation with reasonable performance compared to manual evaluation.
- The system shall be accurate in recognizing handwriting from the answer papers.
- Apart from the initial cost, the system shall be less costly to maintain.
- The system shall be open-source.
- The system shall be usable on any platform.

Chapter 7

Preliminary Design

7.1 Flowcharts

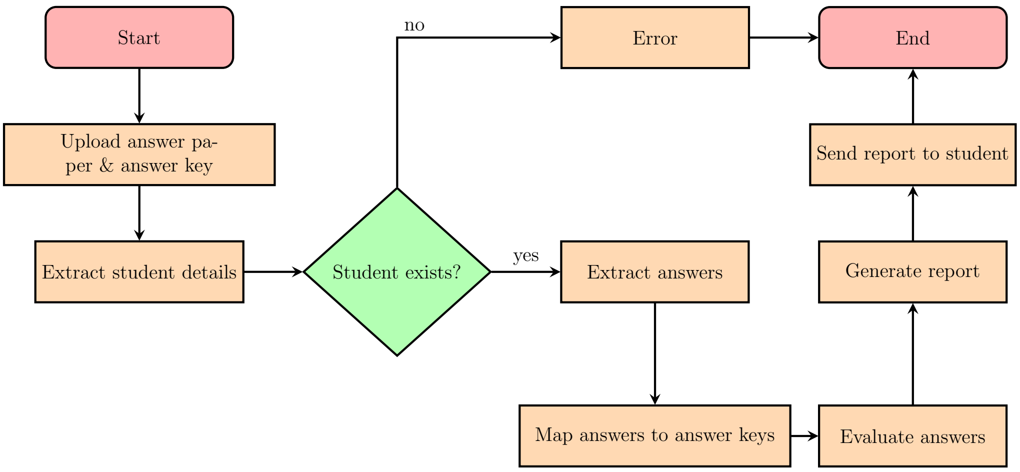

Figure 7.1: Flowchart

The flowchart of our system is shown in Figure 7.1 which represents the work flow. First, the teacher uploads the answer paper and the answer key to the system. The handwriting recognition system extracts the students details from the front sheet of the answer booklet and checks if the student record exists in the database. If no such student exists, there is no further reason to extract the answers from the rest of the answer booklet. So an error is raised. If the student exists, then the answers are extracted and mapped to each question provided in the answer key.

The resultant answer is then provided to the NLP model where semantic analysis is done. Based on the similarity, grades are given and the result is generated and stored in a database. Based on the student details extracted previously, the generated results are sent to individual students.

7.2 Data Flow Diagrams

7.2.1 Level-0 Data Flow Diagram

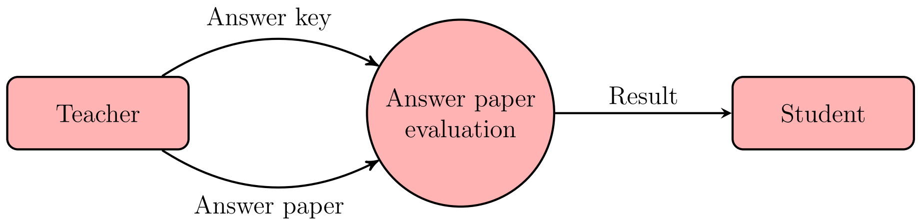

Figure 7.2: Data Flow Diagram Level 0

Figure 7.2 shows the Level-0 Data Flow Diagram of our system. As every level 0 DFD is designed, this diagram also gives the abstract view of our system. The external entities are the *Teacher* and *Student*. Our system is shown as a single process *Answer paper evaluation*. The *Teacher* uploads the answer paper and answer key to the *Answer paper evaluation* process which gives out the result to the *Student*.

7.2.2 Level-1 Data Flow Diagram

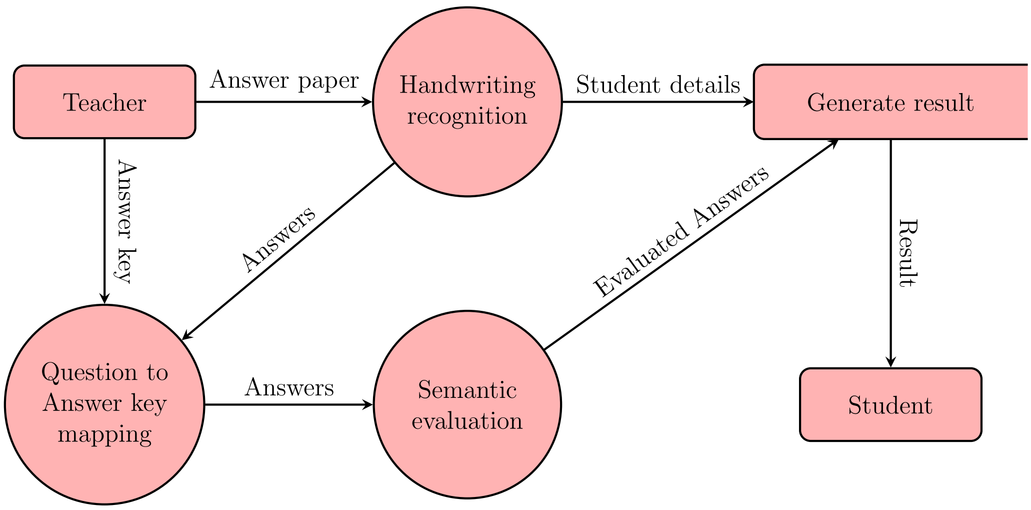

Figure 7.3: Data Flow Diagram Level 1

Figure 7.3 shows the Level-1 Data Flow Diagram of our system. It is a more detailed representation of the Level-0 Data Flow Diagram. The external entities are same as before: *Teacher* and *Student*. The *Answer paper evaluation* process is broken down into several parts. Now, the answer paper from the *Teacher* goes to a *Handwriting recognition* process. The answer key from the

Teacher and the recognized answer from the *Handwriting recognition* goes to a *Question to Answer key mapping* process. The modified answers from the *Question to Answer key mapping* process goes to the *Semantic evaluation* process. Student details from the *Handwriting recognition* process and evaluated answers from the *Semantic evaluation* process goes to the *Generate result* data store where the result is stored and also result to that particular student is sent.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

We presented a method to recognize handwritten texts using a system based on LSTM-RNN model widely applied to transcribe isolated text lines and is inspired from the recent attention-based models. The answers recognized are fed to an NLP model along with the answer key to evaluate the answer paper.

8.2 Future Work

In future, we plan to provide features for sending answer paper copies to student. We also plan to carry out revaluation results. This project can be extended to work with other applications like handwriting recognition of medical prescriptions.

Bibliography

[1] Emmanuel Augustin, Matthieu Carré, Emmanuèle Grosicki, J-M Brodin, Edouard Geoffrois, and Françoise Prêteux. Rimes evaluation campaign for handwritten mail processing. In *International Workshop on Frontiers in Handwriting Recognition (IWFHR'06),*, pages 231–235, 2006.

[2] U-V Marti and Horst Bunke. The iam-database: an english sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5(1):39–46, 2002.

[3] Mario Pechwitz and V Margner. Baseline estimation for arabic handwritten words. In *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition*, pages 479–484. IEEE, 2002.

[4] Sabri A Mahmoud, Irfan Ahmad, Mohammad Alshayeb, Wasfi G Al-Khatib, Mohammad Tanvir Parvez, Gernot A Fink, Volker Märgner, and Haikal El Abed. Khatt: Arabic offline handwritten text database. In *2012 International Conference on Frontiers in Handwriting Recognition*, pages 449–454. IEEE, 2012.

[5] Cheng-Lin Liu, Fei Yin, Da-Han Wang, and Qiu-Feng Wang. Casia online and offline chinese handwriting databases. In *2011 International Conference on Document Analysis and Recognition*, pages 37–41. IEEE, 2011.

[6] Tonghua Su, Tianwen Zhang, and Dejun Guan. Hit-mw dataset for offline chinese handwritten text recognition. 2006.

[7] Tomohisa Matsushita and Masaki Nakagawa. A database of on-line handwritten mixed objects named" kondate". In *2014 14th International Conference on Frontiers in Handwriting Recognition*, pages 369–374. IEEE, 2014.

[8] Masaki Nakagawa and Kaoru Matsumoto. Collection of on-line handwritten japanese character pattern databases and their analyses. *Document Analysis and Recognition*, 7(1):69–81, 2004.

[9] Marcus Liwicki and Horst Bunke. Iam-ondb-an on-line english sentence database acquired from handwritten text on a whiteboard. In *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, pages 956–961. IEEE, 2005.

[10] Lianwen Jin, Yan Gao, Gang Liu, Yunyang Li, and Kai Ding. Scut-couch2009—a comprehensive online unconstrained chinese handwriting database and benchmark evaluation. *International Journal on Document Analysis and Recognition (IJDAR)*, 14(1):53–64, 2011.

[11] Fei Yin, Qiu-Feng Wang, Xu-Yao Zhang, and Cheng-Lin Liu. Icdar 2013 chinese handwriting recognition competition. In *2013 12th International Conference on Document Analysis and Recognition*, pages 1464–1470. IEEE, 2013.

[12] Mark D Shermis and Jill C Burstein. *Automated essay scoring: A cross-disciplinary perspective*. Routledge, 2003.

[13] Y. Zhu, Z. Xie, L. Jin, X. Chen, Y. Huang, and M. Zhang. Scut-ept: New dataset and benchmark for offline chinese text recognition in examination paper. *IEEE Access*, 7:370–382, 2019.

[14] H. K. Janda, A. Pawar, S. Du, and V. Mago. Syntactic, semantic and sentiment analysis: The joint effect on automated essay evaluation. *IEEE Access*, 7:108486–108503, 2019.