Answer Book Valuation Using Semantic Similarity

P. Sunilkumar  
M.Tech Scholar, Department of CSE, Government Engineering College Idukki  
E-mail: sunilkumar.p.kochi@gmail.com

K.S. Sunil  
Assistant Professor, Department of CSE, Government Engineering College Idukki  
E-mail: sunilks@gecidukki.ac.in

Abstract. Conventional examination systems are a tedious task for the universities and the faculties. It takes a lot of time and human effort for the valuation of the answer books, updation of the marks and the declaration of the results. Automation of descriptive answer book assessment would be useful for universities and academic institutions to simplify the valuation system to a large extend. It ensures a uniform valuation and also helps to publish the results without much delay after the examination. We design a system for automatic assessment of descriptive answer books of technical subjects. Semantic similarity is a metric used to assess the similarity between documents and it gives the degree of similarity as a numeric value. It can be used to value the students’ answer books by checking the similarity with original answers given in the answer key and then award appropriate marks.

1. Introduction
In conventional examination systems, expertized faculties are valuing the answer books by comparing with the answers given in the answer key. It requires large amount of time and efforts to complete the examination process and declare results. Since the valuation is a manual process there may be chances to occur human errors during valuation. Automated valuation of descriptive answer book is an innovative idea and it has several advantages over the conventional method of valuation. Time and efforts required for the valuation can be reduced to a large extend and the human errors can be eliminated by using this method and hence it ensures a uniform valuation of the answer books. It is possible to declare the results without much delay after the examination and it is one of the great advantages of this system from the students point of view.

The proposed system will scan answer books and the answer keys and convert them to images. These images will be converted to text format using handwritten word image classification in machine learning. Using semantic similarity methods, the students' answers are compared with the answer keys and assigning appropriate marks.

Daniel Keysers, et al. [3] describes Google's online handwriting recognition system that can support 22 scripts and 97 languages. The important objective of the system is to provide a fast and highly accurate text entry for smart phones as well as touch enabled devices. D. Thenmozhi, et al. [2] proposes a path to find text similarity between a pair of texts by considering various factors. It uses a method to determine set of clauses present in the texts by resolving conjunctions in complex sentences that identify hidden triples from the text. An approach for
finding out the semantic similarity between concepts in Knowledge Graphs such as WordNet and DBpedia is presented in Ganggao Zhu, et al. [4]. It comes up with a semantic similarity method, namely wpath, to combine KG based and corpus based approaches. Ming Liu, et al. [6] proposes an innovative semantic matching method for large documents in the academic and educational clusters. Lexical and semantic similarity measure approach for long answer evaluation is used in Riya Goswami, et al. [8]. This paper aims to introduce a model which programatically evaluates the long answers from the examinee and hence reduce the time and effort of human intervention as well as make the evaluation procedure impartial to the entire user’s. An innovative approach that trains a Fully Convolutional Network (FCN) to predict text line structure in document images is presented in Riya Goswami, et al. [8]. Jija Das Gupta, et al. [5] presents a novel ensemble classifier-based off-line handwritten word recognition system following a holistic approach. Xiangping Wu, et al. [9] proposes a new method for unconstrained offline handwritten word recognition by combining position embeddings with residual networks (ResNets) and bidirectional long short-term memory (BiLSTM) networks. Avani Sakhapara, et al. [1], have designed and implemented a machine learning-based subjective answer grader system (SAGS) using two algorithms, namely latent semantic analysis (LSA) and information gain (IG) for the generation of grades. A hybrid approach for sentence paraphrase identification is discussed in Muhidin Mohamed, et al. [7].

Our work: We proposed a neural network based classifier composed of Convolutional Neural Network (CNN), Recurrent Neural Networks (RNN) and Connectionist Temporal Classification (CTC) layers to recognize text from the handwritten words, which is a unique method for word classification which can give good performance over the above discussed models. We used Universal Sentence Encoder (USE) and Bidirectional Encoder Representations from Transformers (BERT) for obtaining semantic similarity of texts. These two are the state of the art methods being used in the semantic similarity based applications and research. Moreover these two methods are futuristic and will ensure more improvements in future.

2. Preliminaries

Artificial Neural Network is a mathematical model that tries to simulate the structure and functionalities of biological neural networks. A simple mathematical model (function) usually known as an artificial neuron forms the basic building block of every artificial neural network. A Convolutional Neural Network (CNN) is comprised of one or more convolutional and then followed by one or more fully connected layers as in a standard multilayer neural network. The architecture of a CNN is designed to take advantage of the 2D structure of an input image or other 2D input such as a speech signal. Recurrent Neural Networks (RNN) are similar to feed-forward neural network with no limitations regarding backloops. Here the information is transmitted in backward direction also. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior. Long Short Term Memory (LSTM) is one of the recurrent neural networks topologies. In contrast with basic recurrent neural networks, it can learn from its experience to process, classify and predict time series with very long time lags of unknown size between important events. Connectionist Temporal Classification (CTC) is a type of neural network output and associated scoring function, for training RNNs such as LSTM networks to tackle sequence problems where the timing is variable. It can be used for tasks like handwriting recognition or recognizing phonemes in speech audio. The Universal Sentence Encoder (USE) encodes text into high-dimensional vectors. These vectors can be used for natural language tasks like text classification, semantic similarity, clustering etc. Bidirectional Encoder Representations from Transformers (BERT) has been a popular technique in Natural Language Processing (NLP) since Google open sourced it in 2018. BERT is the first deeply bidirectional, unsupervised language representation, pre-trained using only a plain text corpus.
3. System Design
The system which we propose is an automated answer book valuation, which value the scanned answer books and assign appropriate marks using machine learning methodologies for valuing answer books and assigning the scores of the students’. This system calculates the students’ score by checking semantic similarity between students’ answers and original answers. Here the answer books are the scanned images of students’ handwritten answer sheets. Each pages in the answer book will be recorded as separate image files. Filename of the image file will be the combination students’ register number and page number in the answer book. These answers are to be converted into text format for further processing.

The proposed system consists of two modules. Conversion of handwritten images to text format and the valuation of the answer books. In the first module, the answer books are scanned and converted into text format and in the second module we apply semantic similarity methods to find the degree of similarity between the students’ answers and the original answers to assign students’ scores. Figure 1 shows the different steps in the system design.

3.1. Conversion of Answers to text format
Conversion of handwritten images to text is a challenging task because of the huge variations in individual writing styles. The rise of artificial intelligence, along with machine learning and deep learning, are opening up almost limitless possibilities in this area.

In this module, the scanned handwritten answer books are converted to corresponding text format. Machine learning and image processing methodologies are being used for this image to text conversion. Several steps are involved in the conversion process. Figure 2 shows the functional block diagram of handwritten images to text conversion.

- **Annotation:** The co-ordinates of all the words available in the scanned handwritten images will be extracted. It helps to record individual word images from the answer books.
- **Split into Words:** The words in the scanned handwritten images will be copied into separate image files for further processing. It uses the coordinates obtained in the annotation step to identify the bounding boxes of individual words.
- **Prepare Dataset:** IAM handwriting dataset is used for train and test word classification model. It is a popular dataset used for several NLP applications. It contains 115,320 isolated and labeled words.
• Build Model: A neural network classifier will be built to process the dataset and to perform the word classification. The details of the model are described in Section 4.

• Train and Test Model: Training and testing will be done on the model with dataset so as to detect the text format for word images. The details of the step will be described in Section 4.

• Detect Word: Here the system detects the word in text format by inputting the word image into the classification model.

• Combine Words: All the detected words will be combined to form the answer text. Semantic similarity will be done on this answers with the answer key.

3.2. Answer Book Valuation
In this module, the converted answer texts are compared with the original answers using semantic similarity methods and assign the scores of each answers based on the similarity. The original answers in the answer key are also converted to text format using the steps discussed in Section 3.1. The detailed working of this module is given below.

• Preprocessing of the answers: In this step we preprocess the dataset used for semantic similarity. Preprocessing helps to improve the accuracy and efficiency of the similarity checking. The preprocessing include tokenization, removing stop-words, etc. Preprocessing will be done on students’ answers as well as on the answer key.

• Vectorize the answers: Here we convert the answer text to vector format. Pretrained models are used for vectorizing. USE and BERT are used for for vectorizing the text.

• Check similarity: Here we check the semantic similarity between students’ answers and original answers in the answer key. The scores will be calculated with respect to the degree of similarity. Scores of all questions attended by the students are combined to obtain the final score of the student.

4. Implementation
The system is implemented in Python as given in the design. Google Colab and Jupyter Notebook are used as the development environments. This system is implemented using TensorFlow and Keras libraries and the visualizations are done using TensorBoard.

4.1. Handwritten Text Recognition (HTR)
Our system uses a neural network based classifier for recognizing words. We identify the handwritten text by using CNN for feature extraction and RNN with CTC for sequence labeling. This design achieves good accuracy rate and demonstrates the potential for future improvement.

• Preprocessing: The images from the dataset may not be in same size, therefore we resized word images to 128x32 without distortion. We normalized the gray-values of the image which simplifies the task of the neural network. Data augmentation is done by copying the image to random positions or by randomly resizing the images.

• Model Creation: A classification model is built with 5 CNN layers, 2 RNN layers and one CTC layer. CNN layers are used to extract features from the word images. The output feature map has a size of 32x256. The popular LSTM implementation of RNNs is used in this classifier. It has more robust training-characteristics than vanilla RNN.

• Training & Testing: 95% of the dataset is used for training and 5% of the dataset is used for testing the model. 128x32 sized input is supplied to the model. CNN will extract features from the word images. The feature sequence contains 256 features per time-step. The RNN propagates relevant information through this sequence. The output sequence of
the RNN is mapped to a matrix of size 32x80 which is fed into the CTC layer. For loss calculation, we feed both the ground truth text and this matrix in to the operation. The ground truth text is encoded as a sparse tensor. The length of the input sequences must be passed to both CTC operations. We now have all the input data to create the loss operation and the decoding operation. Finally we get the output from the model.

- **Word Recognition**: 128x32 sized input image is supplied to the model which predict the word in text format. Combine these words to form the answers in text format. The detail of classifier model is: *Tensorflow version 1.4.0, Input size: 128 x 32, Training and Testing ratio: 95:5, Batch Size: 50, Epochs: 38, Optimizer: RMSPropOptimizer.*

### 4.2. Check Semantic Similarity

Semantic similarity is computed with students answers and the answer key. Students’ score is computed from the degree of similarity obtained. We use two popular methods USE and BERT for checking the semantic similarity.

- **Convert to text**: We convert the handwritten answer books in to text format.
- **Preprocessing of students’ answers & original answers**: Preprocessing is performed on both students’ answer books and on the answer key.
- **Vectorize students answers and original answers**: Here we convert the answer text in to vector embedding. USE has the vector size of 512 and BERT has the vector size of 768. Vector size is fixed regardless of the size of text.
- **Check similarity**: Here we check the semantic similarity on vectors to find the degree of similarity between students answers and the original answers. Cosine similarity measure is used to find the similarity of vectors. The result of the similarity will be between 0 and 1. A value close to 1 will be treated as most similar answers or correct answers and value close to 0 will be treated as dissimilar answers or incorrect answers. The score will be calculated with respect to the degree of similarity.

![Figure 3. Accuracy - Testing](image1.png) ![Figure 4. Error Rate - Testing](image2.png)

### 5. Conclusion and Result Analysis

**Handwritten text detection**: We implemented a word classifier to recognize words from word images. IAM handwriting database is used as the dataset. We got 65.96% accuracy and 14.73% loss for the word classification. Figure 3 shows the word classification accuracy on testing. X-axis denotes iterations and Y-axis denotes accuracy rate. Figure 4 shows the word classification character error rate on testing. X-axis denotes iterations and Y-axis denotes error rate.

**Answer Book Valuation**: The answer book valuation have conducted using Universal Sentence Encoder (USE) and Bidirectional Encoder Representation for Transformers (BERT). USE showed an overall accuracy of 73.82% when 15 answer books are evaluated while BERT showed 74.53% accuracy for the same set of answers. Table 1 shows the results obtained.
Measuring semantic similarity of text or documents or concepts is a crucial component in many applications. Handwritten text recognition also a complex functionality in NLP. We implemented an answer book valuation system using semantic similarity. It used neural network containing CNN, RNN and CTC for detecting text from word images. We have obtained 65.96% accuracy and 14.73% loss in text detection. More elaborate training and testing on wider database and new technology classifiers may give better results in handwritten text recognition for the future designs. USE and BERT give exceptional performance in text similarity based applications. The automatic valuation methods will replace the conventional valuation by using the advancements in machine learning and natural language processing in near future.

### References

[1] Bhakti Chaudhari Rishabh Gada Aakash Mishra Shweta Bhamshali Avani Sakhapara, Dipti Pawade. Subjective answer grader system based on machine learning. *Springer Conf, Feb-2019*.

[2] Chandrabose Aravindan D. Thenmozhi. Paraphrase identification by using clause-based similarity features and machine translation metrics. *The Computer Journal, Sep-2016*.

[3] Henry A. Rowley Li-Lun Wang Victor Carbune Daniel Keysers, Thomas Deselaers. Multi-language online handwriting recognition. *IEEE Trans., May-2016*.

[4] Carlos A Iglesias Ganggao Zhu. Computing semantic similarity of concepts in knowledge graphs. *IEEE Trans, Sep-2016*.

[5] Bhabatosh Chanda Jija Das Gupta, Soumitra Samanta. Ensemble classifier-based off-line handwritten word recognition system in holistic approach, iet image process. *IEEE Trans., Aug-2018*.

[6] Zepeng Gu Ahmed Zeeshan Ming Liu, Bo Lang. Measuring similarity of academic articles with semantic profile and jointword embedding. *Tsinghua Science and Technology, Dec-2017*.

[7] Mourad Oussalah Muhlidin Mohamed. A hybrid approach for paraphrase identification based on knowledge-enriched semantic heuristics. *Springer, Apr-2019*.

[8] Avik Bisai Alok Ranjan Pal Riya Goswami, Somik Karmakar. A knowledge based approach for long answer evaluation. *IEEE Conf, Jan-2018*.

[9] Jinghan You Yulun Xiao Xiangping Wu, Qingcai Chen. Unconstrained offline handwritten word recognition by position embedding integrated resnets model. *IEEE Signal Processing Letters, Jan-2019*.

### Table 1. Answer Book Valuation using USE and BERT

| Reg.No | Original Marks | USE Marks | Accuracy% | BERT Accuracy% |
|--------|----------------|-----------|-----------|----------------|
| 0001   | 13             | 10        | 78        | 11             |
| 0002   | 10             | 9         | 91        | 11             |
| 0003   | 3              | 7         | 42        | 9              |
| 0004   | 15             | 11        | 75        | 12             |
| 0005   | 11             | 10        | 96        | 12             |
| 0006   | 11             | 13        | 80        | 15             |
| 0007   | 15             | 9         | 63        | 11             |
| 0008   | 19             | 20        | 95        | 20             |
| 0009   | 9              | 11        | 77        | 14             |
| 0010   | 7              | 12        | 56        | 16             |
| 0011   | 17             | 10        | 63        | 15             |
| 0012   | 18             | 12        | 71        | 16             |
| 0013   | 9              | 13        | 67        | 15             |
| 0014   | 6              | 9         | 65        | 10             |
| 0015   | 11             | 8         | 80        | 11             |
| Overall Accuracy | 73.82 | 74.53 |