

# Comparison of the Methods use for Automated Testing of the Exam Papers

Aashka Vijapura

*Department of Computer Science*  
aashkavijapura@gmail.com

Itisha Paul

*Department of Computer Science*  
itishapaul23@gmail.com

Leena Haorongbam

*Department of Computer Science*  
lhaorongbam22@gmail.com

Joshua Paul Gomes

*Department of Computer Science*  
joshuapaulgomes@gmail.com

Dr. Shubha Bhat

*Department of Computer Science*  
subha22@gmail.com

**Abstract**—Manual correction of subjective examination papers can be time-consuming, difficult and may lead to unfair checking. Automation of a descriptive answer assessment procedure will help many universities and academic institutions. This paper aims to compare different approaches of NLP that will help in building the automated assessment mechanism of the examination papers. Some of the various methods discussed here are CNN, POS-Tagger.

The aim is to select the best method to implement a system for automatic assessment of a descriptive answer.

**Index Terms**—NLP, CNN, POS-Tagger,

## I. INTRODUCTION

Academic institutions have descriptive papers for the assessment of the students. It requires a lot of time to correct the paper by the professors. As the volume increases, it becomes time-consuming and there is no alternative solution for this. Hence, the testing platform needs to be automated by a method that is efficient enough to correct the answers written by students in their own words.

Here we are taking the answers typed by the students for the questions that are pre-formed by the teachers. The questions formed by the teachers as well as the answer for those questions given by the teachers will be added in a data set and the answers given by the students will be saved in another data set. The appropriate approach in NLP that we will use shall help us deal with a different tasks that is good in its word and phrases selected from the correct answers and should also omit the probability of selecting the terms not related to the questions.

We are also trying to focus on the approaches which extract the keyword from the answers given by the teachers and categorize them as important keywords, less important keywords or technical keywords. The comparison also focuses on the selection of approaches which are efficient in predicting the synonyms of the actual keyword and also providing antonyms for them in order to figure out wrong answers given by students.

## II. APPROACHES OF NATURAL LANGUAGE PROCESSING

### A. Rules Based Approaches

Rules act as the basis of Rules-Based System, all rules are applied on the input data that is the answers matched from a textbook or given by teachers in our scenario. The purpose of selecting the rule-based system is if we want a highly accurate output, it is difficult for the machine to generate specific rules just by observing a small amount of data in case if the output should be too accurate. Rule-based approach is based on matching the text with manually designed words, part of speech or another way of word representation in a sentence. The text in our case is the answer given by the student. It also excels in extracting the hybrid sequence part of speech and words. One of the main features of Rule-Based System is that we can include human crafted rules. It has four components.

1. List of rules which form the knowledge base.
2. Inference engine, it takes action by studying the interaction of inputs with the rule base.
3. Temporary memory.
4. User interface.

Rule-base approach concentrates more on pattern matching and parsing. Rule-based approach has high performance in specific cases as they have high recall and low precision but their performance degrades when generalized.

### B. Traditional Machine Learning Approach

This approach includes probabilistic modeling, likelihood maximization and linear classifiers. These machine learning approaches are characterized by training data, feature engineering which includes word type, surrounding words, capitalization, plural, etc. It also includes training on parameters that will then do the fitting on test data and finally it will infer the most accurate keywords which in our scenario involves finding the most probable keywords from solution given by teachers. Some of the methods that we use as traditional machine learning methods are Support Vector Machine(SVM) and Bayesian networks. Machine Learning can be used for the task of Tokenization by which it will break the answers

given by teachers into phrases and words that machine can understand. It can also be used for sorting out the type of keywords we have extracted from the answers and it can also cluster the keywords which are referred to the same context.

Machine learning can also be used for unsupervised NLP where it will find the similarities between the items. One such method is a matrix factorization method for example "I ate the apple lying on the table" this method will associate the word ate with apple. Traditional machine learning approaches are good for sequence labeling. They can also improve the neural network methods that we use for NLP.

### *C. Neural Network Approach For NLP*

This method is quite similar to traditional machine learning methods. These are the few differences between Neural Network Approach For NLP and traditional machine learning methods. These are the few differences between Neural Network Approach For NLP.

- The most important characteristic of traditional machine learning in NLP is feature engineering which will be skipped in neural networks because networks learn the important features on its own.
- Streams of raw parameters are given to the neural networks without engineering features.
- The input to the neural networks is large training corpus.

Some of the Neural network Methods used for NLP are RNN and CNN. Machine learning methods just make a prediction based on past observation while deep learning approaches do not only predict but also focus on correctly representing the data such that it is suitable for our prediction. Major concepts in a neural network for NLP is using an embedding layer where when we embed the words that transform from being distinct words into mathematical objects that can be operated on. Convolutional network is used for sentence classification. recurrent neural networks along with sequence to sequence learning can be used for machine translation. End to End Memory Networks can be used for question answering tasks. Sequence to sequence models are mainly used in question answering systems, chatbots and machine translation.

## **III. RULE BASED ALGORITHM IN NATURAL LANGUAGE PROCESSING**

### *A. Transformation Based Learning*

The basic idea behind this method is to define the minimum number of rules so that the system can achieve the highest performance on the application of those rules sequentially. This algorithm can be used with other statistical methods like part of speech tagging and natural language understanding.

Transformation Based Learning can change one state to another using transformation rules in order to find an appropriate tag for each word. The biggest advantage of Transformation Based Learning is that it allows a user to have linguistic knowledge in a readable form. The outcome of Transformation Based Learning is properly ordered sequence of transformation where a word represented by a particular tag is transformed into another represented by a different tag in the given context.

For example, Jackson was playing on the ground. A typical Transformation Based Learning tagger will identify Jackson as a noun if a given rule is mentioned in prior.

Transformation Based Learning usually begins with some simple solution to the problem and then it will go through many cycles. In each cycle, it will execute a transformation that gives the most benefit and it will stop at a cycle where no further benefits can be given. The biggest advantage of this algorithm is simple rules that are sufficient for tagging, the complexity of tagging is reduced by human-generated rules. Some of the disadvantages are it does not provide tag probabilities and training times are usually very long especially on large data sets.

### *B. Regular Expression*

The Regular expression is a sequence of characters that define a search pattern. These types of expression can be used by searching algorithms to find and replace strings in a valid input. Each character in a regular expression is a meta character with a special meaning or a regular character. These meta characters and regular characters can be used to identify a text of a given pattern or instances of it. Regular expression matching is used to tag words. `RegexTagger` is a subclass of `SequentialBackoffTagger`, it can be positioned before a `DefaultTagger` so as to tag words that n-gram tagger missed. Patterns are saved in `RegexTagger` it then iterates over patterns by calling `chooseTag()`, it then returns the first expression tag which matches the current word. Hence if two given expressions get matched then tags of the first one will be returned without even trying the second expression.

## **IV. TRADITIONAL MACHINE LEARNING APPROACH FOR NLP**

### *A. Support Vector Machine (SVM)*

Support Vector Machine is a vector-spaced based machine learning method where the main objective is to find the decision boundary between groups in the training data that has maximum distance from any point. If our data set contains different tags, then SVM works by taking the data points and output the hyperplane that best separates the tag. This hyperplane is called the Decision Boundary, anything that falls on one side will be classified of one type other than that falls on the other side of the hyperplane. To find the best hyperplane the one whose distance is to the nearest element in each tag is largest is chosen. To apply SVM in Natural Language Processing for text classification, we treat a text as a bag of words and every word that appear in a bag has a feature. The value given to that feature will be the frequency of the word used in the answer sheet. Every word should be treated as a vector with multiple dimension, each vector represented in the frequency of one of the words used in the text. Choose the labeled text and convert them to vectors using word frequencies and feed them into the algorithm to produce a model, so when we have a new unlabeled text we want to classify we can convert it into vector feed it into the model

that will output the tag of the text. Depending on the answer we are checking we can do the following things.

- collapse part numbers into a single token.
- up weighting, which counts a word as if it has occurred twice.
- feature selection, which selects those words which are specific to a particular class and drops the other words. .

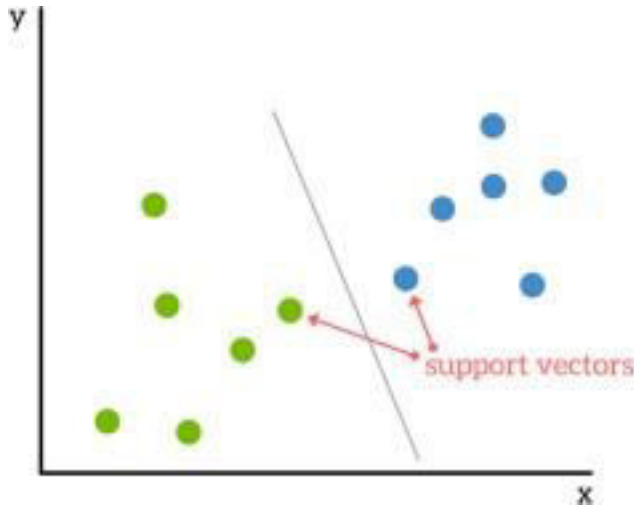


Fig. 1. Support Vector Machine

### B. Naive Bayes Classifier

This is a simple classification rule based on Bayes rule. It relies on representing a document in a simple representation called a bag of words. Positive and negative inputs are given to the model and we have a bag of positive words and negative words then we may count how many times that word appears in the document, in order to classify that document as a positive or negative document. In our scenario, we can classify answers related to a particular topic.

Naive Bayes Classifier is mostly used in sentiment analysis, spam filtering, recommendation systems, etc. They have a disadvantage where predictors have to be fully dependent even if they are quick and easy to implement. The predictors are conditional in reality and this hinders the overall efficiency of the classifier.

## V. NEURAL NETWORK-BASED APPROACH FOR NLP

### A. Convolutional Neural Networks (CNN)

CNN consists of one or more complex layers, continually followed by a subsampling step and one or more fully connected layers. The framework of CNN is arranged by taking the 2D structure of an image or speech signal. CNN is trainable and has fewer metrics.

A CNN, alternatively accompanied by interconnected layers, comprises of a number of convolutional and subsampling layers. The input to a convolutional layer is  $m \times m \times r$  image

where  $m$  is the height and width of the  $m$ =image and  $r$  is the number of channels.

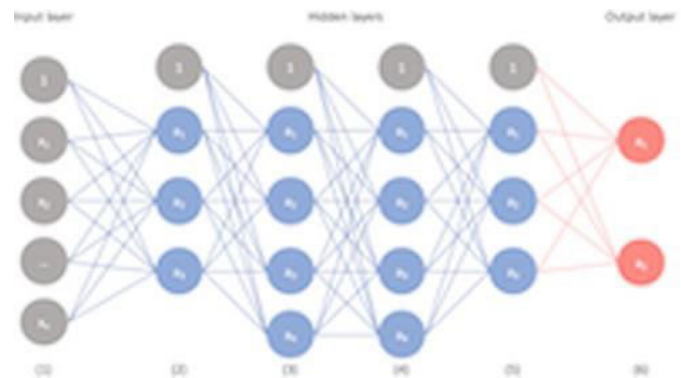


Fig. 2. Convolutional Neural Network.

The most natural fit for CNN is a classification tasks like sentiment analysis, spam detection, top categorization. Convolutions lose information about local order of word,  $m$  so POS tagging or entity extraction is a bit harder to fit into a pure CNN architecture. CNN achieves very good performance across data sets and new state-of-the-art on a few. The input layer is a sentence made of concatenated word2vec word embedding followed by a convolutional with multiple filters, followed by a max pulling layer. Finally, a softmax classifier.

### B. Reinforcement Learning

Reinforcement learning refers to goal-oriented algorithms, which learn how to attain a complex objective or maximize the points won in a game over many moves. They can start from a blank slate, and under the rights conditions, they achieve superhuman performance. Reinforcement learning tackles the complicated issues of equating concrete actions with the prolonged results they actually generate.

Reinforcement learning can be interpreted utilizing the concepts of agents, environment states, action and rewards.

- Agent: An Agent takes actions.
- Action: Action is the set of all possible moves the agent can make. Action is almost self-explanatory, but it should be noted that agents usually choose from a list of discrete, possible actions.
- Environment: The world the agent is passing in and that needs respond to the agent.
- State: A State is a concrete and immediate situation in which the agent finds itself i.e. a specific place and moment, an instantaneous configuration that puts the agent in relation to other significant things.
- Rewards: A Reward is a feedback by which we measure the success or failure of an agent's actions in a given state.

Reinforcement learning varies by supervised and unsupervised learning.

- Unsupervised Learning: The unsupervised learning describes a challenging situation in which for every observation  $i=1,2,3,n$ , we observe a vector of measurements  $x(i)$

but no associated response  $y(i)$ . It is not possible to fit a linear regression model since there is no response variable to predict. This situation is referred to as unsupervised. An Example includes Clustering.

- **Supervised Learning:** In the supervised learning domain, for each observation of a predictable measurement of  $x(i)$ , where  $i=1,2,\dots,n$ , there is an associated response  $y(i)$ . We wish to fit a model that relates the response to the predictor with the aim of accurately predicting the response for future observations or better understanding the relationship between the response and the predictors. Some Examples are Linear Regression, Logistic Regression etc.

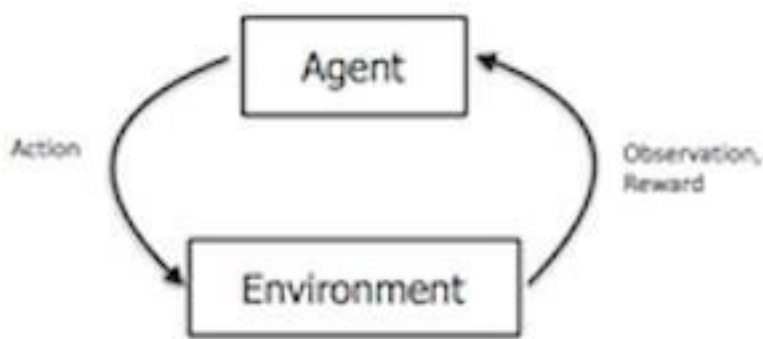


Fig. 3. Reinforcement Learning

Reinforcement Learning is a general-purpose framework for decision making. It is an agent with the capacity to act and each action will influence its future state. The success of the agent is measured by a scalar reward signal. Agents have to select actions to maximize future rewards. During the process of text classification of data from the answers given by teachers, we can create an environment and an agent. The agent will try to classify the text in an arbitrary manner. Based on the result, the agent will get a reward due to which it will decide next action. It will do so by going through areas of neural networks. Agent must explore possible states of actions which will help to classify the answer in a rightful manner.

## VI. COMPARING DIFFERENT NLP METHODS

The traditional and rule-based approach is good for sequence labeling. Some ideas in neural networks are very similar to traditional and rule-based approaches like word2vec are similar in concept to distributional semantic methods. We can use methods from traditional approaches to improve neural network approaches. The neural network approach has many applications and a lot of research in NLP. The inability to construct long-distance dependencies is one of the flaws of convolutional neural networks(CNN) which is required in many NLP tasks. CNN is effective in the context they can extract interpretative clues from contextual information. It finds difficulty in preserving sequential order and building long-distance contextual details. RNN is efficient in the way

it can memories the result from the previous computation and used the result of the previous computation in current computation. On comparing RNN and CNN model, both can have equal efficiency and in some NLP tasks, RNN can work better than CNN. Both neural network and non-neural network approaches can be useful for contemporary NLP in their own right, they can also be used or studied in detail for maximum potential benefit.

## VII. PART OF SPEECH TAGGING

Part of Speech(POS) Tagger is used to tag part of speech such as nouns, adjectives, etc to words in a sentence by reading each of the sentences. Then these tags of the words are parsed using the grammar rules.

Part of Speech	Tag
Noun	n
Verb	v
Adjective	a
Adverb	r

Fig. 4. Part of Speech Tags

There are neuro and rule-based hybrid tagger with high accuracy and require less training data. To increase the accuracy of the tagging system we can increase the training data or improve the quality of the corpus used for training. The hybrid tagger mentioned uses multi-layer perceptrons, it suffers from the problem of non-convergent when increasing the training data so to overcome this we can apply  $M^3$  neural networks.

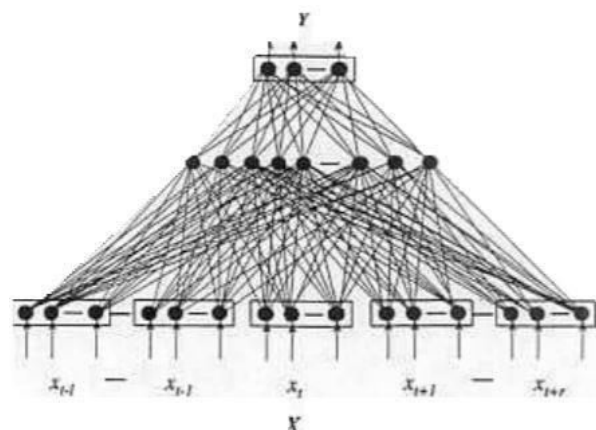


Fig. 5. Neuro Tagger.

The hybrid system has two main elements, first is a neuro tagger which is used as initial-state annotator and second is a



rule-based corrector that corrects the output of neuro tagger. At first, the sentence is fed to a neuro tagger which initializes a tag to it, the rule-based corrector is then used to refine the output given by neuro tagger and presents the end results of tagging. The neuro tagger consists of three perceptron layers. Neuro tagger cannot obtain rules when a conditional part is obtained from lexical information, so a rule-based corrector rectifies this fault of neuro tagger.

#### VIII. CLASSIFYING THE TEXT PRESENT IN THE ANSWER SHEET CONTAINS THE FOLLOWING STEPS.

- Take each of the sentences.
- Perform word segmentation.
- Perform POS(Part of Tagging).
- Extract the keywords based on keywords extracted from the actual solution.
- Include the synonym based on the semantic relation between words and compare these to keywords to credit score to the student.

Classify the message based on the most suitable methods from the different methods mentioned above.

#### IX. CONCLUSION

Examination plays a very important role in academic institutions and most of the academic institutions go for subjective examinations, though there are exams which include only objective questions or combination of both objective and subjective question, objective question plays a very crucial part in checking a person's aptitude skills but they fail to check the depth of conceptual and theoretical knowledge of the person. The paper discusses different approaches to evaluate subjective papers. Comparing different techniques we have come to the conclusion that neural networks are the best approach for our model to evaluate descriptive answer sheets. Marks are awarded to the student when the keywords in the answer sheet match with the keywords extracted from the answer source or with the synonym of the keyword. This system can thus solve the problem of checking paper in bundles and miscorrection by awarding marks fairly to each answer sheet.

#### REFERENCES

- [1] Siyu Hu, Hao Zhang, Yin Chen. Realization of Test Paper Score Cumulating System Based on Digital Image Processing, International Journal of Engineering Science and Computing, March 2017.
- [2] Arnau Ramisa, Fei Yan, Francesc Moreno-Noguer, and Krystian Mikolajczyk. BreakingNews: Article Annotation by Image and Text Processing, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.
- [3] Nilima Sandip Gite, "Implementation of Descriptive Examination and Assessment System", International Journal of Advance Research in Science and Engineering, vol. 7, no. 7, March 2018.
- [4] Bhagyashri Kaiche, Samiksha Kalan, Sneha More and Lekha Shelukar, "Online Descriptive Examination and Assessment System", International Journal of Emerging Technology and Advanced Engineering, vol. 4, no. 3, March 2014.
- [5] Aditi Tulaskar, Aishwarya Thengal, Kamlesh Koyande. Subjective Answer Evaluation System. International Journal of Engineering Science and Computing, April 2017.
- [6] Wang Hairui, Wang Hua. Research and Implementation of Multi-agent Based Test Paper Generation Algorithm. International Conference on Computer Science and Software Engineering 2008.
- [7] Amarjeet Kaur, M Sasikumar, Shikha Nema, Sanjay Pawar, "Algorithm for Automatic Evaluation of Single Sentence Descriptive Answer", International Journal of Inventive Engineering and Sciences, vol. 1, no. 9, August 2013.
- [8] MAO Qing-hua. (2003) Application of Artificial Neural Network in Aerospace Experiment Effectiveness Evaluation. Automation of Theory Applications.
- [9] Chua L., Yang L. (1988) Cellular neural network: Theory and Applications. IEEE Transactions on Circuits Systems I, vol. 35, no. 10. pp. 1273 – 1290.