Chapter 15
Study on Sentence and Question Formation Using Deep Learning Techniques

N. Venkateswaran
Department of Management Studies, Panimalar Engineering College, India

R. Vidhya
Department of Computer Science and Engineering, Sri Krishna College of Technology, India

Darshana A. Naik
Ramaiah Institute of Technology, India

T. F. Michael Raj
Department of Computer Applications, SCLAS, SIMATS University (Deemed), India

Neha Munjal
Department of Physics, Lovely Professional University, India

Sampath Boopathi
https://orcid.org/0000-0002-2065-6539
Mechanical Engineering, Muthayammal Engineering College, India

ABSTRACT

Natural language techniques require less personal information to communicate between computers and people. Generative models can create text for machine translation, summarization, and captioning without the need for dataset labelling. Markov chains and hidden Markov models can also be employed. A language model that can produce sentences word by word was created using RNNs (recurrent neural networks), LSTMs (long short-term memory model), and GRUs (gated recurrent unit). The suggested method compares RNN, LSTM, and GRU networks to see which produces the most realistic text and how training loss varies with iterations. Cloze questions feature alternative responses with distractors, whereas open-cloze questions include instructive phrases with one or more gaps. This chapter provides two novel ways to generate distractors for computer-aided exams that are simple and dependable.

DOI: 10.4018/978-1-6684-6782-4.ch015

1. INTRODUCTION

Software for text prediction was developed to help persons who write slowly and to improve communication. It merely uses a few early text fragments to predict the previous phrase that is likely to continue. Currently used methods use a text prediction algorithm to choose the optimal word depending on the current phrase. Artificial Neural Networks, which are machine learning algorithms, are a subset of deep learning. They are constrained, nonetheless, in terms of creating appropriate sentence structures for lengthy sequences. Recently, it has been discovered that deep learning approaches are frequently employed and produce successful outcomes. One of the main factors contributing to their success is their flexibility in choosing the architecture. Machine learning models make judgments based on what they have learnt from the data, whereas neural networks put up algorithms to make decisions on their own that are dependable. Deep learning models can manage more data and anticipate more accurately than machine learning algorithms, producing outcomes that are more precise than those produced by current system technologies(Abujar et al., 2019; Raza et al., 2019).

With categories for voice tagging, word meaning disambiguation, and named entity identification, NLP is a hot issue in academia. An Indo-Aryan language that is descended from Sanskrit, Language is an Indo-Aryan language. For example, Language, the world's 23rd most common language, is processed and pre-processed using methods from Hindi, Sanskrit, Arabic, and English. Many sentences are built using a subject, object, and verb in that sequence. There are two Vachans: single and plural(Ahmad et al., 2020; Sharif et al., 2020). The Objective Nominative Case is expressed by a series of words, phrases, and sentences, with various words standing in for one term, separate terms for one sentence, and different letters for one word. It's derived from a set of letters known as "kakko."

"Sentences are composed of many letters and words."

Language text processing problems can be difficult due to ambiguity, phrase complexity, language grammatical construction, text translation errors, and the difficulty of finding reliable data for text processing algorithms. A language model is a set of estimates made by a self-supervised learning system. Labels are ingested in the data, and by comprehending the specifics of the predicted corpora, researchers may apply transfer learning to enhance the performance of text classifiers. The Indic NLP library offers a general answer to the problems that Indian languages encounter due to their many similarities in terms of writing, phonetics, grammar, and other areas(Al-Aswadi et al., 2020).

Complexities in the Reginal Language: Each noun in the gender-classified Reginal Language designates one of three gender kinds. The past, future, and present tenses are the three that are employed in language processing. There are two varieties of vachans: singular and plural. If a word is only represented in the singular, it cannot be processed or stated in the plural. Language is a regional tongue with five main dialects, each of which conveys a unique meaning based on the environment, society, and community. Due to their numerous quasi-words, these words are known as multiquasi words and can lead to ambiguity in utterances. Different Language Resources: WordNet is a connected lexical positioning system that incorporates social lexical memory ideas from psycholinguistics. Adjectives, verbs, and nouns are all equivalent in English and each expresses a fundamental lexical idea. A group of synonyms are connected by several connections. It is one of the 22 recognised languages of India, and the linguistic WordNet was produced using a Hindi language extension approach. Wikipedia is a web resource having nouns, verbs, lemmas, and other categories represented. Architecture of WordNet is shown in Figure 1.