of up to 2.8 bi-lingual evaluation understudy (BLEU) scores compared to various neural machine translation systems. It outperformed the commonly used MT system on a WMT 14 dataset.

Fan et al. [41] introduced a gradient-based neural architecture search algorithm that automatically finds architecture with better performance than a transformer, conventional NMT models. They tested their model on WMT14 (English-German Translation), IWSLT14 (German-English translation), and WMT18 (Finnish-to-English translation) and achieved 30.1, 36.1, and 26.4 BLEU points, which shows better performance than Transformer baselines.

Wiese et al. [150] introduced a deep learning approach based on domain adaptation techniques for handling biomedical question answering tasks. Their model revealed the state-of-the-art performance on biomedical question answers, and the model outperformed the state-of-the-art methods in domains.

Seunghak et al. [158] designed a Memory-Augmented-Machine-Comprehension-Network (MAMCN) to handle dependencies faced in reading comprehension. The model achieved state-of-the-art performance on document-level using TriviaQA and QUASAR-T datasets, and paragraph-level using SQuAD datasets.

Xie et al. [154] proposed a neural architecture where candidate answers and their representation learning are constituent centric, guided by a parse tree. Under this architecture, the search space of candidate answers is reduced while preserving the hierarchical, syntactic, and compositional structure among constituents. Using SQuAD, the model delivers state-of-the-art performance.

4.2 State-of-the-art models in NLP

Rationalist approach or symbolic approach assumes that a crucial part of the knowledge in the human mind is not derived by the senses but is firm in advance, probably by genetic inheritance. Noam Chomsky was the strongest advocate of this approach. It was believed that machines can be made to function like the human brain by giving some fundamental knowledge and reasoning mechanism linguistics knowledge is directly encoded in rule or other forms of representation. This helps the automatic process of natural languages [92]. Statistical and machine learning entail evolution of algorithms that allow a program to infer patterns. An iterative process is used to characterize a given algorithm's underlying algorithm that is optimized by a numerical measure that characterizes numerical parameters and learning phase. Machine-learning models can be predominantly categorized as either generative or discriminative. Generative methods can generate synthetic data because of which they create rich models of probability distributions. Discriminative methods are more functional and have right estimating posterior probabilities and are based on observations. Srihari [129] explains the different generative models as one with a resemblance that is used to spot an unknown speaker's language and would bid the deep knowledge of numerous languages to perform the match. Discriminative methods rely on a less knowledge-intensive approach and using distinction between languages. Whereas generative models can become troublesome when many features are used and discriminative models allow use of more features [38]. Few of the examples of discriminative methods are Logistic regression and conditional random fields (CRFs), generative methods are Naive Bayes classifiers and hidden Markov models (HMMs).

(a) Naive Bayes Classifiers