Survey on Machine Translation Techniques

Sunil Sangve[1], Om Gunjal[2], Lavesh jaykar[3], Saurav Garje[4], Samir Aghav[5], Saurabh Gunge[6]

[1,2,3,4,5,6] *JSPM's Rajarshi Shahu College of Engineering, Pune-411033, Maharashtra, India.*

[1]sunilsangve@jspmrscoe.edu.in
[2]om.s.gunjal@gmail.com
[3]laveshjaykar223@gmail.com
[4]sauravgarje88@gmail.com
[5]samiraghav24@gmail.com
[6]saurabhgunge20@gmail.com

Abstract— **Language is a vital part of human evolution. Being able to convey our thoughts, express our emotions, maintain data and knowledge over a long period of time and interact with other people is all possible because of Language. But, over the years, in different geographical regions, different kinds of languages were developed. This discrimination of language becomes a barrier for people to interact with people outside their geographical zone. Also, learning a different language is a critical and time-consuming process. But machines can be quick to learn the language translation and help us to resolve this issue. In this survey paper, we have summarized various research papers that exercise different Machine Translation (MT) approach which has motivated numerous advancements in the Machine Translation domain with remarkable performance in catching long term dependencies and context of the sentence.**

This review article presents numerous methodologies, approaches, and developments in the subject of Neural Machine Translation in order to inspire future research in this area.

Keywords— **Machine learning, Natural Language Processing, Neural Machine Translation, Machine Translation, Artificial intelligence.**

I. INTRODUCTION

The importance of translation in our daily lives has a different aspect to it than we realize. Language is much more than just a means of communicating words; it is also a means of expressing culture, society, and belief. So, even though English is a widely spoken language all around the globe, but, until you speak what his/her heart speaks (their native language), you won't be communicating effectively and thus, it may lead to some miscommunication. English might not remain the world's most popular language as advancement in Developing countries has resulted in many other languages to grow in importance as well.

So, translation is necessary so that we can make accommodations. These languages are a worldwide economic force. If we are able to achieve proper communication with different cultures, then we will be able to share new and important knowledge as well as information with each other. This might help in increasing the global medical approach as well as different scientific developments from this cultural fusion.

Machine translation is impacting many major industries and domains in this modern world. Various applications of these techniques are:

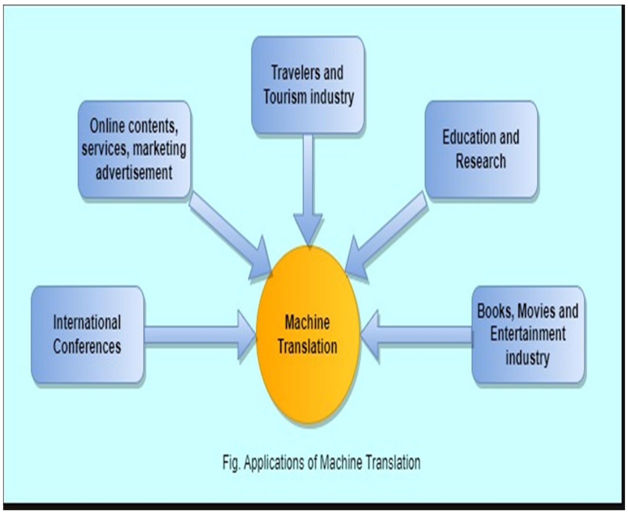

Fig 1. Applications of Machine Translation

II. RELATED WORK

Many researchers have applied different machine translation techniques to translate the given text into the languages we want and much research is still going on to find more effective and efficient techniques for machine translation. Some of those are listed below:

Paper[1] studies byte-level sub-words, notably a BPE technique used at the level of bytes, known as BBPE. It is more effective than using pure bytes and is compacter than character vocabulary.

1

Paper[2] discusses the idea to use an English-to-local language translator for ease of communication between local Nigerian residents.

Paper[3] focuses on developing a hybrid neural network approach to develop a more robust English to French machine translation model.

Paper[4] encourages the representation of the need for conducting translation and transliteration of qualitative research.

Paper[5] aims to reduce training time by reducing precision and large batch training with no effect on accuracy.

Paper[6] offers an upgraded BERT by using average pooling (AP-BERT), to improve the BERT model's capability to learn phrase-level semantic information.

Paper[7] introduces MAML, which is a model-independent algorithm for meta-learning, to handle the problem raised due to low-resource availability in machine translation.

Paper[8] translates missing words from the data. Generates and extends dictionaries for any language pairs.

Paper[9] develops a real-time translator for better communication and to reduce the inability of dictionaries and human translators.

Paper[10] introduces Bidirectional Encoder Representations from Transformers widely known as BERT, which is established as a novel method for language representation.

Paper[11] is intended to inform the client regarding the quality of the machine translation output during testing.

Paper[12] suggests constructing a single neural network for maximizing interpretation performance by making fine adjustments periodically.

Paper[13] suggests an automated translation learning technique which has the capability to solve many of the drawbacks observed in conventional phrase-based translation systems.

Paper[14] presents OpenNMT, an open-source neural MT toolbox (NMT). The system places a premium on efficiency, modularity and extensibility.

Paper[15] suggests that, with proper tweaking and execution, lower precision along with big batch training may reduce training time by around five times on a computer with 8 GPUs.

Paper[16] proposes a new context-aware neural machine translation model to capture extra sentential information which is mostly ignored by other approaches by isolating the sentences.

Paper[17] proposes a denoising auto-encoder which is used for pre-training of a sequence-to-sequence translation model.

Paper[18] proposes a simple network architecture called Transformer.

Paper[19] demonstrates the significant performance gains of using multilingual denoising pre-training over different machine translation tasks.

Paper[20] teaches how to translate only each language and has access to massive monolingual corpora. columns.

III. LITERATURE REVIEW

A study and review of these researches is carried out using a comprehensive survey method and observations are compiled in the observation table.

3.1 Methodology

The survey method starts with analyzing the requirements and searching for Research papers in standard journals and resources such as Springer, Google Scholer, Papers with Codes, IEEE, etc.

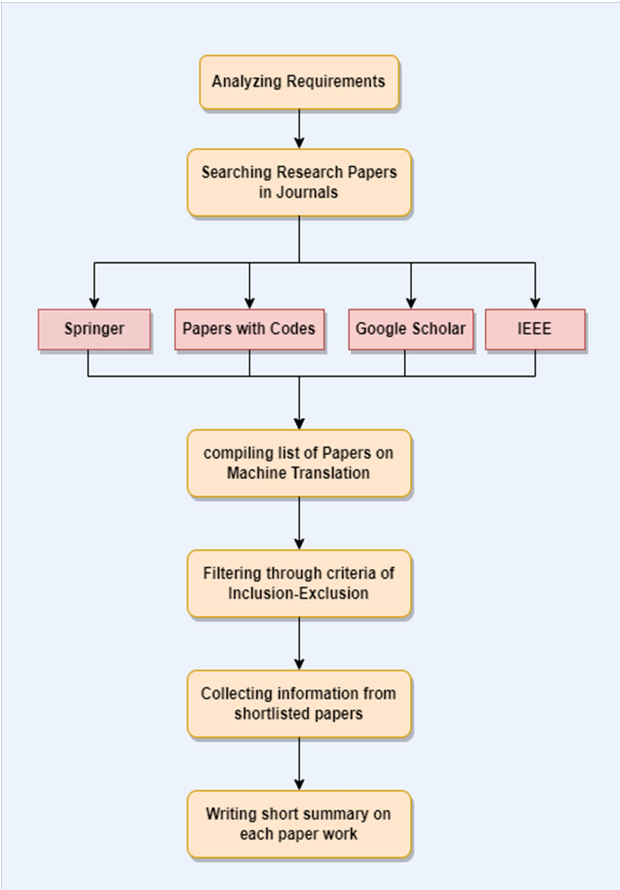

Fig.The Methodology used for survey

Fig. 2. The methodology used for survey

2

All the research works related to Machine translation is compiled together. A specific inclusion-exclusion criteria is defined and all the papers are filtered through that criterion. Then a comprehensive study of shortlisted papers is done and observations are listed in the observation table.

3.1 Observation Table

TABLE I

OBSERVATIONS ON PROPOSED SOLUTIONS, RESULTS AND LIMITATIONS

P. ID	Year	Proposed Solution	Technique and Results	Limitations
P1	2020	Author Changhan Wang et al., proposed the idea of contextualizing BPE to BBPE and using either convolutional or recurrent layers for the implementation.	Contextualizing BPE vocabularies to BBPE (Byte level BPE) vocabularies 'subwords' not only results in short tokenized sentences but also provide a good performance as when we use BBPE on Noisy Character sets, Character-rich languages and Many-to-En Translation the resulting performance is comparable with BPE even with such short size of training dataset.	However, these byte-level subwords may sometimes produce erroneous byte sequences, i.e. sequences that on decoding do not make sense in character sequence.
P2	July 2018	Author Iwara I. Arikpo et al., suggested using a hybrid of transfer and corpus-based models, Object-Oriented methodology for designing and developing a Java based machine translation for local residents to overcome the language barrier for Nigerian residents.	The Application developed was able to produce good results for simple English sentences and thus proving the system useful for translation to language have less magnitude of difference in grammatical structure.	The proposed system was unable to translate complex English sentences and also cannot translate to languages with more magnitude of difference in grammatical structure.
P3	January 2019	Venkata Sai Rishita Middi et al., recommended building a deep neural network for translation from English to French. In this paper, the author also describes different RNN model architectures and step by step manipulation of data to achieve the final translation.	On implementation the proposed model translated English to French with an accuracy of 96.71%.	In end-to-end models, the cause of failure is unknown which then demands modification in model parameters, architecture or even training dataset. Also, the end-to-end model requires a significant amount of training data.
P4	1 March 2010	Author Krishna Regmi et al., discuss the importance of having qualitative research translated and transliterated in different languages for cross-cultural research.	This paper has not only suggested strategies for translation and transliteration but also different techniques to overcome difficulties in those strategies.	Having ancient medical research from different cultures translated into a common language can widen the horizons for medical advancements. A newbie in a qualitative field will face more difficulty in translating since it requires a good understanding of research in both languages.
P5	10 Oct 2014	Author Akshay Suresh Deshpande et al., suggested developing a prototype that uses speech processing hardware for providing a real time translation to users.	The author also discussed the three Automatic speech translation technologies including voice recognition technology, translation technology, and speech synthesis technology in the other person's language. The proposed Voice to Voice Translation System will overcome the language barrier and real life hassles faced by unschooled people.	A proper translation by the proposed system can't be done in noisy areas, while real time communication does not always occur in quiet places
P6	19 March	Shuai Zhao et al., presented an upgraded BERT to improve the	An average pooling technique is applied in this model in contemplation of performing	It has been able to enhance only four tasks: Text categorization,

3

	2022	capacity of BERT model to handle semantic information at phrase level, using average pooling.	token embedding which reconstructs the input. It finally results in significant improvement in the impact of BERT applications in Chinese natural language processing.	recognized named entities, reading comprehension, and summary creation that too for Chinese Language only underlining its less scope for application.
P7	25 Aug 2018.	Author Jiatao Gu et al., introduces MAML, which is a model-independent algorithm for meta-learning, to handle the problem raised due to low-resource availability in machine translation.	It demonstrates that meta-learning frequently outperforms multilingual transfer learning. Only if the input and output spaces are shared across all source and target processes then meta-learning can be used for machine translation with low-resources.	More research should be focused on building multiple meta-models so that A new language has freedom to select a model for adaptation.
P8	17 Sep 2013	Author Thomas Mikolov et al., suggested a translation system that translates missing words or phrases from data. Models which are used in the system are the Skip-gram and CBOW models. It can be trained on a large dataset. Duplicate sentences and special characters were removed and numeric values were also replaced by single tokens.	It can be trained on a large They performed data cleaning and preprocessing on WMT11 datasets and built monolingual datasets for English, Spanish and Czech languages. They translated data using translation matrices and found dictionary errors.	This work can be used to improve dictionaries and phrase tables for any language. But, it cannot be used for low resource language corpora.
P9	Dec 2020	Author Yusrah Bablani et al., gave a brief review of machine translations systems. Machine translation is none other than a part of Natural Language Processing and Artificial Intellect.	A rule-based method known as RBMT, neural network based NBMT technique, statistics-based translation methods SBMT are various techniques used to fully automate machine translation. Essentially, the authors of this research discussed various ways to machine translation for data extraction.	Accuracy is not offered by machine translation on a consistency basis. It cannot concentrate on context information due to various rules and weakness about interpretation and lexical data.
P10	11 October 2018	Author Jacob Devlin et al., introduces Bidirectional Encoder Representations from Transformers widely known as BERT, which is established as a novel method for language representation.	A fine-tuned pre-trained BERT model with only one extra output layer produces cutting-edge models for a variety of NLP applications, including question answering and language inference. BERT is both theoretically and experimentally simple.	The BERT's key disadvantage is the computing resources required to train and fine-tune the model. In order to make BERT models suitable for production, knowledge distillation, quantization, pruning, and other approaches must be used.
P11	1 Sep 2020	Author Marina Fomicheva et al., gave the idea about quality estimation (QE) in a machine translation system. Quality estimate attempts to assess the quality of translated information in the absence of a reference translation.	They gave an unsupervised approach to quality estimation in which machine translation does not require any training. In the absence of a reference translation, quality estimation seeks to measure the quality of translated content.	But, since it is an unsupervised approach to quality estimation, it works as black box and it is really hard to explain the results.
P12	Sep 2014	Author Dzmitry Bahdanau et al., aimed to make a single neural network which can maximize translation performance.	Encoder-decoder approach is best for neural machine translation (NMT). Novel architecture is used in the issue of fixed-length context vectors causing problems in translating long sentences. They used RNN search model on English to French translation which outperforms the encoder-decoder model. This model is a promising step towards the future as compared to available phrase-based machine translation.	It cannot handle rare words or words belonging to rare languages. That is one of the crucial limitations of this method.
P13	26 Sep 2016	Author Yonghui Wu et al., is an automated translation learning technique which has the capability to solve many of the drawbacks observed in conventional phrase-	In comparison with Google's phrase-based production system, it decreases translation errors by 60% on average and establishes a well-balanced relation between the efficiency of word-separated delimited	The results achieved were generally poor when compared with the effort required to set up such implementations.

4

		based translation systems.	models and the flexibility of character-separated models.	
P14	10 Jan 2017	Author Guillaume Klein et al., introduces OpenNMT which is an open-source neural machine learning toolbox (NMT).	The system promotes efficiency, flexibility, and extensibility in order to assist NMT research while remaining competitive.	The toolbox includes modeling and translation assistance, but for new research and novel, innovative techniques one can't find everything on that toolkit, they need to do it from scratch.
P15	1 June 2018	Author Myle Ott et al., proposes that, with proper tweaking and execution, lower precision along with big batch training may reduce training time by around five times on a computer with 8 GPUs.	It managed to achieve 43.2 BLEU on the WMT '14 English-French dataset, in 8 and half hours on 128 GPUs and also obtained a new cutting-edge technology 29.3 BLEU after training for 1 hour and 25 minutes on 128 GPUs.	This scaling only improves time requirements for training but doesn't affect translation quality directly.
P16	25 May 2018	Author Elena Voita et al., proposes a novel model for neural machine translation which is aware of context, to capture extra sentential information.	Such information is mostly ignored by other approaches but this method measures the relationship between induced attention distributions and coreference relations to implicitly capture anaphora.	In many cases its effectiveness is not very significant, increasing the BLEU score by just 0.7 or 0.6 points.
P17	2020	Author Mike Lewis et al., proposes a denoising auto-encoder which is used for pre-training of a sequence-to-sequence translation model.	With only target language pretraining, it excels a back-translation system by BLEU score of 1.1.	Although it increases the quality of translation, the change is not very significant.
P18	2017	Author Ashish Vaswani et al., introduced the powerful 'Transformer' architecture for Machine Translation.	It improves the performance and takes significantly less time for training because of the parallelization and self-attention mechanism.	Self-attention allows a specific window size to attend to. Hence, more advanced techniques of Multi-headed Attention models with bidirectional encoder are preferred over it.
P19	22 Jan 2020	Author Yinhan Liu et al., proposes mBART which is a newly developed denoising text in several languages for pre-training entire seq-to-seq models.	Pre-training enables it to be fine-tuned directly, increasing performance for low resource machine translation by a BLEU score of 12 points whereas for document-level and unsupervised models by more than 5 BLEU points.	This pre-training process needs high resource settings.
P20	2018	Author Guillaume Lample, et al., tells how to translate if only each language has access to a big monolingual corpus.	In some languages, machine translation systems attain near-human-level performance, and it surpasses the state-of-the-art model by more than 11 BLEU points even when no parallel sentences are present.	But they require a very large monolingual corpora, without it these models won't give significant accuracy score.

IV. ADVANTAGES

- Human Translators take a tremendous amount of time to translate the books, documents, articles, online content, etc. But machine translators can translate huge volumes of data within a relatively short time.
- It takes a few years for a person to learn a new language. But machines can learn new languages a way earlier than any human.
- A human translator can have a good hold over a maximum of 5 to 6 languages. But multilingual machine translators can translate one sentence into hundreds or thousands of languages quickly.

V. CHALLENGES

After reviewing all these research papers, we identified certain difficulties or challenges faced by researchers working in this domain. Following are some of the most significant challenges:

- Translating the languages having Limited Resources available.
- Building Multilingual Translators to translate a sentence into all languages in the world.
- Achieving Real-Time Translation Speed.
- Providing Human-Like Translation.

VI. CONCLUSION

The research reflects the interest of tech-giants like Google, Amazon, Netflix, etc. along with several tech-startups in the field of machine translation.

We can clearly see the evolution of technology in the domain of machine Translation from this paper. From traditional Rule based and statistical translation systems up to the neural machine translators, transformers models powered by attention mechanisms have infinitely long reference window.

It underlines the fact that even after a lot of advancement in this domain, there is a large scope in translation of regional languages with limited resources. There are more than 5000 languages in the world and most of them don't have enough data available for training machine translation models. This scarcity of resources is driving a new dimension of the machine translation techniques which uses hallucinated machine generated images related to a sentence and then try to translate it into another language through the intermediate hallucinated image.

The Speech recognition and synthesis techniques are also contributing in providing machine translations in audio form. Different types of such new research frontiers will grow this exciting and crucial field of machine translation in future.

These translation techniques can be broadly categorized in following types:

TABLE II
COMMON CATEGORIES OF MACHINE TRANSLATION TECHNIQUES

Sr. No.	Summary of Machine Translation Techniques	
	Name of technique	**Idea**
1	Rule Based Machine Translation (RBMT)	This approach runs a grammatical examination of the input and target languages to produce the translated sentence. This is a complex and time-consuming approach.
2	Statistical Machine Translation (SMT)	For translating, this method depends upon statistical models. These models are unable to understand the sentence's context.
3	Hybrid Machine Translation (HMT)	This is approach is composed of both statistical and rule-based translation systems. But, significant amount of proofreading is required by human.
4	Neural Machine Translation (NMT)	This translation process exploits the deep learning neural network-based models to make the translations. They are good at capturing the context better.
5	Visual Hallucination based Machine Translation.	This method converts a sentence into intermediate visual representation and using this visual representation the sentence is translated into another language.

References

[1] Prof. Wang, C., Cho, K., & Gu, J. (2020). Neural Machine Translation with Byte-Level Subwords. Proceedings of the AAAI Conference on Artificial Intelligence, 34(05), 9154-9160.

[2] Iwara I. Arikpo, Iniobong Dickson, Development of an automated English-to-local-language translator using Natural Language Processing, International Journal of Scientific and Engineering Research 9(7):378-383, July 2018

[3] Venkata Sai Rishita Middi, Middi Appala Raju, Tanvir Ahmed Harris," Machine translation using natural language processing", January 2019

[4] Prof. Krishana Regmi, "Understanding the process of translation and transliteration in qualitative research", International journal of qualitative methods, 1 March 2010

[5] Assi. Prof. Akshita H. Utgikar, "Voice to Voice language translation system", International journal of engineering research and technology (IJERT), Vol.3 10 Oct 2014

[6] Shuai Zhao, Tianyu Zhang, Man Hu, Wen Chang, Fucheng You, "AP-BERT: enhanced pre-trained model through average pooling",Springer 19 March 2022", June 2013

[7] Jiatao Gu, Yong Wang, Yun Chen, Kyunghyun Cho, Victor O.K. Li "Meta-Learning for Low-Resource Neural Machine Translation", arXiv:1808.08437 [cs.CL], EMNLP 2018, 25 Aug 2018.

[8] Tomas Mikolov," Exploring similarities among languages for machine translation", arXiv: Computation and language,17 Sep 2013

[9] Yusrah Balbani," Survey on Text-to-Text Machine Translation",2nd International Conference on Computational Sciences and Technologies, 17-19 December 2020 (INCCST'20), MUET Jamshoro

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" 11 October 2018, arXiv:1810.04805 [cs.CL]

[11] Marina Fomicheva," Unsupervised Quality Estimation for Neural Machine Translation", Transactions of the Association for Computational Linguistics (2020) 8: 539–555,1 Sep 2020

[12] Dzmitry Bahdanau," Neural Machine Translation by Jointly Learning to Align and Translate", arXiv: Computational and Linguistics, Sep 2014

[13] Yonghui Wu," Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation",arXiv: Computation and Language, (2016),26 Sep 2016

[14] Giullaume Klein," OpenNMT: Neural Machine Translation Toolkit", ACl,10 Jan 2017

[15] Sergey Edunov,David Grangier," Scaling neural machine translation", arXiv:cs:computation and language,, vol 1,1 June 2018

[16] Elena Voita,"Aware Neural Machine Translation Learns Anaphora Resolution ",arXiv:1805.10163v1 [cs.CL],25 May 2018

[17] Mike Lewis," Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension"

,10.18653/v1/2020.acl-main.703,58th Annual Meeting of the Association for Computational Linguistics, 2020

[18] Ashish Vaswani," Attention Is All You Need", Advances in Neural Information Processing Systems 30 (NIPS 2017)

[19] Yinhan Liu,"Multilingual Denoising Pre-training for Neural Machine Translation", arXiv:2001.08210v2 [cs.CL] ,22 Jan 2020

[20] Guillaume Lample, Myle Ott," Phrase-based and neural unsupervised machine translation",2018 conference on empirical methods in natural language processing,5039-5049,2018