JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

MACHINE TRANSLATION TECHNIQUES AND CHALLENGE'S: A STUDY

Dr. G. Nagalakshmi
Assistant Professor
Computer Science
National Sanskrit University: Tirupati

Abstract: Translation play a major role in communication. It acts as an intermediate between source language and target language. Technically Machine Translation involves the same process of conversion of text or sentence form of one language to another language. There is no human interaction in this text translation from one language to another; the conversion is being handled by a machine. Machine translation is an application of computers under Artificial Intelligence towards the task of translation of texts from one natural language to another. Several approaches have been adopted in the recent times for developing the machine translation system. Each has its own advantage and disadvantages. The machine translation systems use primarily different techniques: neural, statistical, and rule-based. A common approach is rule-based, which combines language and grammar with the use of dictionaries. There are many challenging aspects of machine translation like the large variety of languages, alphabets and grammars, the task to translate a sequence (a sentence for example) to a sequence is harder for a computer than working with numbers only, there is no one correct answer (e.g.: translating from a language without gender-dependent pronouns, he and she can be the same). This paper majorly concentrates on various machine learning techniques and its challenges.

Key words: -Machine Translation, Statistical Machine Translation, Artificial Intelligence, Rule-based MT, NMT, Comparison, Techniques, Example-based Machine Translation, Challenges

I. INTRODUCTION:

Machine Translation is the use of computer software and algorithms that automatically translate text or speech from one language to another, with no human input. MT is a part of the subfield called computational linguistics, AI, and NLP and has been developed and grown dramatically within the past few decades. It ensures to make languages accessible to everyone since they do not speak each other's language, for that matter, across different cultures and regions. MT has been defined as the research of how to use the aid of computers to translate words of one language into another, conceived even before the first computer, was developed in 1947 by Warren Weaver one year after its invention. Since then, MT has been considered to be one of the most challenging tasks in the field of NLP.

How Machine Translation Works

Modern MT systems rely on various algorithms to process and translate language. Here's an overview of how they generally function:

- **Preprocessing**: Input text is cleaned and standardized, removing unnecessary characters or formatting that could hinder translation.

- **Tokenization**: The text is broken down into smaller units like words or subword segments.
- **Encoding**: The input text is transformed into a format that the model can understand, often using word embeddings or neural networks.
- **Translation Process**: The MT model applies learned rules, patterns, or statistical relationships to generate a translation.
- **Post-processing**: The translated text is refined for grammatical accuracy and fluency, making it resemble human-level output.

II. TYPES OF MACHINE TRANSLATION TECHNIQUES

Based on its methods of functioning, machine translation can be categorized into three main approaches: rule-based machine translation, statistical machine translation, and neural machine translation. The three approaches use varied techniques and methodologies to translate words.

Rule-based machine translation: In rule-based machine translation, explicit linguistic rules and dictionaries are relied upon to translate words. The rules are created manually by linguistic experts, such as syntactic and grammatical patterns, vocabulary mappings, and transformation rules. RBMT systems process the source text, apply predefined rules, and generate the corresponding translation in the target language. Some advantages of RBMT are control over grammar and syntax, and thus it is suitable for some domains and languages. However, RBMT is not good for idiomatic expressions, language variations, and extensive lexical resources.

Statistical machine translation: This type of machine translation uses statistical models and algorithms to translate text. SMT systems have been trained on large parallel corpora made up of aligned sentences within the source and target languages. Models can learn statistical patterns and relationships between the source and target languages while learning. In the process, translation will be done on the most probable translation to be generated based on these statistical patterns learned. SMT was the dominant approach before the rise of NMT and permitted the effective handling of complex language structures and variations. However, SMT often requires extensive linguistic resources and suffers from capturing long-range dependencies.

Neural machine translation: Neural machine translation represents a paradigm shift in machine translation, using artificial neural networks to model the translation process. NMT models, particularly the transformer-based ones, are very popular because of their impressive performance. These models consist of encoder and decoder parts, and attention mechanisms which allow for the capturing of contextual dependencies. The ability of NMT models is to learn end-to-end generation of translations without making explicit rule-based or statistical alignments. The benefits that NMT has is being able to handle long-range dependencies, capture context, and output fluent translations.

III. IMPORTANCE OF MACHINE TRANSLATION:

MT has the capability to bridge language barriers and enable communication across different languages. It is becoming very important in our interconnected world for several reasons:

Global Communication: MT allows people of diverse native language speakers to communicate; people can converse, negotiate with one another, or carry out joint ventures all across the globe without speaking even a word of multiple other languages.

Breaking Down Language Barriers: MT systems remove the language barrier, translating text, documents, websites, and other content into multiple languages, so more people can access information. It is particularly important to businesses, governments, and organizations that want to reach a global market or audience.

Efficiency and Cost-effectiveness: Human translation by this method can be very time-consuming and costly, particularly when a large volume of material is involved. MT is quicker and cheaper and allows for volumes of text to be translated rapidly.

Multilingual Content Creation: MT allows the production of multilingual content in e-commerce, marketing, and entertainment sectors. Companies can easily and speedily produce localised content for various markets with high customer engagement.

Support for Low-resource Languages: Even though such high-resource languages as English, Spanish, and Chinese have such developed MT models, even this is helping to bridge support for less-spatial or digitally present speakers. This has opened channels for information to be able to reach the minority or indigenous language speakers.

Integration with Other Technologies: It is integrated with other technologies, such as voice recognition, chatbots, and artificial intelligence systems. This helps in enriching the user experience through applications such as virtual assistants, online customer support, and real-time language interpretation.

Cultural Exchange and Knowledge Sharing: MT will help break the barrier of languages so that there could be better cultural exchange and sharing of knowledge through which people would get content in multiple languages, including research papers, books, and media, from other cultures.

Real-time Translation: With the advent of MT, voice conversations, video calls, and live events can be translated in real time. So, immediate communication is made possible across language divides.

Enhancing Education: MT can make academic resources available to students, educators, and researchers in languages they otherwise might not understand, furthering learning opportunities and the development of academic collaboration across borders.

Improvement Through Machine Learning: Improvement through machine learning and large-scale data training is an improvement process that continuously improves the accuracy and fluency of MT systems. Such progress is very essential for producing high-quality translations in contexts ranging from formal legal documents to casual conversations.

Business: Machine translation enables the organization to translate business documents, contracts, marketing materials, product descriptions, and customer support communications. That way, an organization is able to communicate better with international markets.

Research and collaboration: For the academic and research communities, machine translation has emerged as an indispensable tool that enables collaboration and sharing of knowledge. Through the medium of translation, scientists in various countries with differing linguistic background can read scientific articles, conference proceedings, and results of research findings. Machine translation helps spread research globally by accelerating scientific advancement and encouraging multidisciplinary cooperation.

Travel and tourism: Machine translation is an important tool for a traveller and tourist because it provides instant translations of signs, menus, directions, and other information. It enables a person to navigate unknown places, communicate with the people, and understand their culture. Machine translation applications and devices make travel easy and accessible, and people can go to new destinations with confidence.

IV. MACHINE TRANSLATION APPLICATIONS AND PLATFORMS

Translation software has been designed to speed up the translation process. For this reason, machine translation software is most suited to companies that require one-off quick translations where the degree of accuracy is not essential.

Google Translate: Google Translate is one of the most popular and widely used machine translation platforms. It supports translation between a large number of languages and offers various features such as text translation,

Website translation, document translation, and real-time speech translation. Google Translate is available as a web application and mobile app, making it accessible across different devices.

Microsoft Translator: Microsoft Translator is another highly used machine translation service. It is a multi-language translation platform that can translate text, documents, websites, and speech. The applications available with it range from the web-based translator to mobile apps, API services, and integration with Microsoft Office and other products from Microsoft.

DeepL: DeepL is a popular machine translation platform known for its quality translations. It uses neural machine translation technology to give accurate and fluent translations. DeepL supports translation between various language pairs and provides features such as document translation, web page translation, and integration with other applications through its API.

Amazon Translate: Amazon Translate Amazon Web Services (AWS) provides translation services in its product by offering automatic text translation between multiple languages, from easy content localization and real-time chat translation to translating in customer support. Its key feature is providing the most competent translations using neural machine translation models.

Yandex.Translate is a machine translation service provided by Yandex, a Russian technology company. It supports translation between many language pairs and offers text translation, document translation, and website translation. Yandex.Translate is available as a web-based application and provides additional features like dictionary look-up, pronunciation assistance, and language detection.

SYSTRAN: SYSTRAN is a machine translation software service provider. It provides business translation solutions to organizations, such as neural machine translation models, customizable translation engines, and translation API services. SYSTRAN caters to different industries, such as e-commerce, customer support, and localization.

V. MACHINE TRANSLATION CHALLENGES:

It delves into various aspects of machine translation that make the task complex in terms of achieving accurate and contextually relevant translations. One of the major challenges is capturing and utilizing contextual information from the source text. Context is not just about the immediate surrounding words but also about the broader discourse and cultural background. This is still challenging in adapting a machine translation system to specific domains such as legal, medical, or technical. Since models are trained on general data, they may not perform optimally when they face domain-specific terminology. and hence requires extra efforts for adaptation to the domain.

Another significant challenge is to keep the grammatical correctness and syntactic coherence. Various languages have different structures for syntax and grammar, so it is challenging to create a translation that follows the grammatical rules of the target language and maintains the intended meaning of the source text. Even with the advancements made in neural networks and neural machine translation models, sometimes sentences generated are grammatically incorrect or awkwardly phrased, especially when the target language has complex sentence structures.

Accurate translation of technical, scientific, or domain-specific terms is yet another challenge. These terms often have specific meanings and connotations within their respective fields, requiring a deep understanding of the subject matter to generate accurate translations.

Many machine translation systems perform well for widely spoken languages because they have abundant training data for that language. However, they often struggle with low-resource languages that lack an extensive bilingual

diaspora. This is challenging because creating good translation systems or models for such languages requires innovative approaches to data scarcity.

Machine translation can break the language barrier, promote multicultural understanding, and facilitate global cooperation. It opens up new opportunities for communication, information exchange, and cooperation across diverse linguistic communities. Therefore, achieving contextually appropriate translations in machine translation systems is a complex task that requires addressing various challenges. Capturing and utilizing contextual information, handling idiomatic expressions and cultural nuances, maintaining grammatical correctness and syntactic coherence, and accurately translating technical or domain-specific terms are the areas that require attention. It can help pave the way for effective cross-lingual communication and be used in the advancement of global collaboration and understanding by addressing these challenges.

VI. CONCLUSION:

Machine translation has for decades been an active area of artificial intelligence research. MT is a hard problem, since natural languages are highly complex, many words have various meanings and different possible translations, sentences might have various readings, and the relationships between linguistic entities are often vague This study uniquely focuses on the techniques and challenges of capturing of machine translation, by presenting a detailed examination of the different machine translation techniques or systems handle idiomatic expressions and metaphors, distinguishing between these types by highlighting the need for improved contextual understanding. This paper addresses the very critical issue of translating domain-specific terminology, with particular emphasis on domain adaptation and integrated specialized knowledge, so as to ensure improvement in translation accuracy. The suggestion from this study is that future research needs to work toward the development of models that better capture long-range dependencies and relationships between sentences to enhance the contextual understanding by applying those techniques. In addition, it also details the challenges machine translation has on different fields.

VII. REFERENCE:

[1] Shei, C.C., 2002. Teaching MT Through Pre-editing: Three Case Studies. Proceedings of the 6th EAMT Workshop: Teaching Machine Translation. Manchester, England: European Association for Machine Translation, pp.89–98. Available from: https://aclanthology.org/ 2002.eamt-1.10.

[2] Somers, H., 2003. Machine translation in the classroom. In: H. Somers, ed. Computers and Translation: A translator's guide. John Benjamins Publishing Company, Benjamins Translation Library, vol. 35, pp.319–340. Available from: https://doi.org/10.1075/btl.35. 20som.

[3] Tarasenko, R.O., Amelina, S.M. and Azaryan, A.A., 2019. Features of the use of cloud-based translation systems in the process of forming information competence of translators. CTE Workshop Proceedings, 6, p.322–335. Available from: https://doi.org/10.55056/cte.394.

[4] Tarasenko, R.O., Amelina, S.M. and Azaryan, A.A., 2020. Improving the content of training future translators in the aspect of studying modern CAT tools. CTE Workshop Proceedings, 7, p.360–375. Available from: https://doi.org/10.55056/cte.365.

[5]. Dictionary Based Word Translation in CLIR Using Cohesion Method Mallamma V Reddy1 and M. Hanumanthappa Department of Computer Science and Applications, Bangalore University, Bangalore, INDIA

[6]. Singla Savita and Baghla Seema, " Hybrid Approach for English to Punjabi Translation System for News Paper Headlines in a Specific Domain" , International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 11, November – 2013.

[7] Kaur Harjinder and Laxmi Vijay , "A Web Based English to Punjabi MT System for News Headlines ", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 6, June 2013 www.ijera.com

[8] Gaule Monika and Josan S. Josan, "Machine Translation of Idioms from English to Hindi", International Journal Of Computational Engineering Research , Vol. 2 Issue. 6 , Oct 2012.

[9]. Retrieved from https:// sites.google.com/site/ profrmksinha / research-projects/machine tran

[10]. Josan S. Gurpreet, Lehal S. Gurpreet, "A Punjabi to Hindi Machine Translation System" , Coling 2008: Companion volume – Posters and Demonstrations, pages 157–160 Manchester, August 2008.

[11]. Rani Sumita, Laxmi Vijay, " Direct Machine Translation System from Punjabi to Hindi for Newspapers headlines Domain", International journal for Computer and Technology, Vol 8, No 3, June 3 0 , 2 0 1 3. [12]. Retrieved from en.wikipedia.org/wiki/Interlingual http:// _ machine_translation

[12] R.M.K. Sinha, A. Jain "AnglaHindi: English to Hindi Machine-Aided Translation System". International Journal of Computer application Vol. 3, No. 3, Oct 2008.

[13] Sitender, Seema Bawa "Survey of Indian Machine Translation Systems". IJCST Vol. 3, Issue 1, Jan. - March 2012 ISSN : 0976-8491 (Online) | ISSN : 2229 4333 (Print).

[14] Mishara Vimal, Mishara RB. "Study of Example Based English to Sanskrit Machine Translation" department of computer engineering .Institute of technology, Banaras Hindu University, Varanasi India.

[15] Khaled Shaalan "Rule-based Approach in Arabic Natural Language Processing" International Journal on Information and Communication Technologies, Vol. 3, No. 3, June 2010.