DOI 10.47176/kurmanj.6.4.1

Meta Analysis

SRPH

KURMANJ, 2024; 6(4): 1-5

Machine Translation Evaluation Research in Persian-English Language Pair: A Non-Statistical Meta-Analysis

Fereidoon Sadeghzadeh Yazdi[1,*], Mohammad Reza Hashemi[2]

[1,2] Department of English Language and Literature, Ferdowsi University of Mashhad, Iran.

***Corresponding Author:**
sadeghizadeh@fum.ac.ir

Received: 13 July, 2024
Revised: 27 November, 2024
Accepted: 18 December, 2024
Published: 25 December, 2024

ABSTRACT

It seems an established fact that machine translation has proved its unique utility capacities for global communication. However, evaluating the quality and performance of MT has been and still is a challenge, especially in Persian-English language pairs. Against this background this article examines the current state of research on machine translation evaluation of such pairs. The study reviewed research on evaluating Persian-English MT published since the so-called "neural turn", searching academic databases using keywords, using a coding framework based on strengths, weaknesses, challenges, and limitations of MT systems, and categorizing information like identified weaknesses and proposed improvements. The results reported in this study include: a) a lack of systematic research due to limited industrial developers in Iran, b) academic disinterest, and c) a need for tailored metrics. Finally, the results will be discussed and suggestions will be provided and areas for future research will be mentioned.

Keywords: Translation accuracy, Translation adequacy, Neural machine translation, Machine translation evaluation

Introduction

Background of the Study

Machine translation (MT) is currently widely used for communication purposes (e.g. social media) across the world. While professional human translation is considered by many translation studies scholars to provide the best quality in most of the cases, it is time-consuming and expensive. Combining human and machine translation can be a promising strategy [1]. The quality of machine translation has improved significantly in recent years, leading to the need for continuous evaluation of its performance.

According to the scholars in the field, there are two main approaches to machine translation: rule-based and corpus-based [1,5,6]. Rule-based machine translation relies on linguistic rules programmed into the computer, teaching it how to translate based on linguistic knowledge. On the other hand, corpus-based machine translation trains programs using large collections of bilingual texts to find sentence equivalents. Statistical approaches use probability theory to determine the likelihood of sentence pairs being translations of each other.

Neural machine translation, which implements multiple layers of linear models and non-linearity, has become the dominant approach since 2015. Therefore, previous evaluations of machine translation systems do not seem to be scientifically applicable to these revolutionized systems which carry the same name and logo but utilize revolutionized algorithms and systems.

Machine translation can be categorized into three applications: assimilation, dissemination, and communication [6]. Assimilation involves translating foreign text to understand its content, which is the most

Publisher: Scientific Research Publishing House (SRPH), Shirvan, Iran
http://srpub.org
Email: kurmanj@srpub.org

common use of machine translation. Dissemination refers to translation for publication, where accuracy and adequacy are crucial, often requiring human post-editing. Communication includes live chats, emails, and social media conversations [6,7].

The quality and time restrictions vary for each of the applications. Post-editing is necessary to adjust the output of machine translation to fit the linguistic context, target language structures and norms, target language culture, and target language polysystems [8,9]. While these tasks are challenging for machine translation, human post-editing is currently essential. However, implementing machine translation still offers significant time and cost savings, emphasizing the importance of understanding hybrid human-machine translation [3,8,10].

Persian-English MT evaluations

The evaluation and review of previous MT evaluations of Persian-English language pair is essential for several reasons. First, it can assist in identifying the strengths and weaknesses of existing MT systems. This information can then be used to improve the design and development of new MT systems. Second, it can help to identify the challenges and limitations of MT in general. This information can then be used to develop better MT evaluation methods. Third, it can help to raise awareness of the limitations of MT and ensure that users of MT systems are aware of the potential for errors.

A number of previous MT evaluations of Persian-English language pair have employed a variety of methods to assess the quality of MT output. Some common methods are as follows: 1) Human evaluation: Human evaluators read and assess the MT output and then compare it to the original text. This is the most reliable method of MT evaluation, but it is also the most time-consuming and expensive. 2) Automatic evaluation: Automatic evaluation methods use statistical techniques to measure the similarity between the MT output and the original text. These methods are faster and cheaper than human evaluation, but they are not as reliable. 3) Hybrid evaluation: Hybrid evaluation methods combine human evaluation with automatic evaluation. This can be a more reliable and efficient way to evaluate MT output.

This article examines the findings of previous MT evaluations of Persian-English pair, mostly the ones conducted after the neural-network turn in 2017 [4].

Methods

Selection Criteria

To ensure comprehensive coverage of relevant research on the evaluation of English-Persian machine translation (MT) systems within the Iranian context, a systematic search strategy was employed. This included searching academic databases like Web of Science, Scopus, and Google Scholar using targeted keywords such as "machine translation," "English-Persian," "MT evaluation," "Iran," and specific evaluation metrics (e.g., BLEU, METEOR). The search was further extended beyond peer-reviewed journals and conference proceedings to capture valuable insights from registered dissertations and theses deposited in credible digital repositories -e.g. Iranian Research Institute for Information Science and Technology (IranDoc). These repositories typically undergo rigorous vetting processes to ensure the quality and legitimacy of deposited materials.

Studies were included if they met the following criteria: a) Published or deposited since the "neural turn" (2017-2024) to capture the most recent trends and advancements in MT evaluation. b) Clearly focused on the evaluation of MT systems, specifically addressing the performance metrics of those systems with respect to the English-Persian language pair. c) Published in peer-reviewed journals, credible conference proceedings, or deposited in recognized digital repositories that adhere to established quality control standards.

This approach was taken to ensure the inclusion of relevant studies from diverse sources, providing a well-rounded understanding of the current state of MT evaluation in the chosen context. Ultimately, due to the limited research in this area, 18 relevant studies were selected. study's aim is to get metal flow and distributions of equivalent stress on some special sections such as longitudinal and transverse sections under processing tube tension-reducing.

Data Analysis

Bibliographic Information

During the data collection process, bibliographic information for each included study was extracted, including:

- Author(s)
- Year of publication/deposit
- Title
- Publication type (e.g., journal article, conference paper, dissertation)
- Source (e.g., journal name, repository name)
- DOI (if available) ight-node hexahedral element type is taken, at the same time 2280 elements and

Qualitative Analysis and Synthesis

Coding Framework:

to systematically analyze the extracted data from the 18 studies, a thematic coding framework was developed.

[DOI: 10.47176/kurmanj.6.4.1]

This framework consisted of pre-defined categories and sub-categories capturing various aspects of MT evaluation in the context of English-Persian language pairs. These categories included:

- The strengths and weaknesses of existing MT systems for Persian-English translations.
  - Specific strengths identified (e.g., fluency, accuracy in specific domains)
  - Specific weaknesses identified (e.g., difficulty with specific language features, limited domain adaptation)
- The challenges and limitations of MT in general.
  - Specific challenges identified (e.g., handling idiomatic expressions, cultural nuances)
  - General limitations discussed (e.g., reliance on training data, lack of human-like understanding of context)
- The development of better MT evaluation methods.
  - Proposed or analyzed evaluation methods (e.g., new metrics, human evaluation approaches)
  - Strengths and weaknesses of different evaluation methods discussed
- The awareness of the limitations of MT.
  - Strategies or recommendations for users to be aware of MT limitations discussed (e.g., user education, appropriate application scenarios)

Coding Process:
Each study was carefully reviewed, and relevant information was extracted and coded using the developed framework. This process involved systematically assigning codes to data segments based on their thematic relevance.
Data Analysis:
Once the coding was complete, the coded data was analyzed to identify recurring themes and patterns across the studies. This involved techniques like thematic analysis, where coded segments were compared and contrasted to identify broader themes and sub-themes within each category.

Results

An overview of the findings of the studies

The selected research seemed to be virtually unanimous in some of their findings. Overall, the overwhelming majority of the studies pointed out some of the most notable strengths and weaknesses of MT.
As for the strengths, it can be said that existing Mt systems:
a) can be used to translate large amounts of text quickly and easily.
b) can be accessed from anywhere in the world,
c) are constantly being updated with new data, which can improve the quality of the translations.
As for the weaknesses of existing MT systems the following are worth mentioning:
a) They are in some cases inaccurate, especially when translating complex or idiomatic language.
b) They produce unnatural and/or incorrect translations.
Moreover, the studies predominantly mentioned main challenges and limitations of MT in general. The former can be listed as: a) It is difficult to capture the nuances of human language. b) It is difficult to translate cultural references. c) MT systems are still not as accurate as human translators. The latter can be put as:
a) Existing MT evaluation methods are often based on word-level metrics, which can be misleading.
b) There is a lack of awareness of the limitations of MT among users.
Furthermore, the selected studies discussed the need for better MT evaluation methods. Existing MT evaluation methods are often based on word-level metrics, such as BLEU and METEOR [10]. These metrics can be useful for measuring the fluency of a translation, but they do not always correlate with the accuracy of the translation. There is a need for MT evaluation methods to be more comprehensive and to take into account the accuracy, fluency, and naturalness of the translation. These methods should also be capable of measuring whether or not an MT system can recognize the cultural nuances of the source language.
The selected studies have also addressed the importance of user awareness of the limitations of MT as follows:

- It is important for users of MT systems to be aware of the limitations of MT.
- MT systems are not perfect, and they can often produce inaccurate or unnatural translations.
- Users of MT systems should always proofread the translations carefully before using them.
- Users of MT systems should also be aware of the cultural limitations of MT.

A non-statistical meta-analysis

T Despite the fact that the selected studies have contributed to the improvement of the field, few points of consideration seem notable:
First, there is an overt lack of systematic research in the field of machine translation evaluation in Persian-English language pair due to factors including the following: First, there are a limited number of industrial machine translation developers in Iran: There are only two MT systems developed in Iran: "targoman.ir" and "faraazin.ir". Their efforts are concentrated more on naturalness and accuracy of Persian language part [11,12]. However, due to the limited budget and apathy in developing MTs in Iran, their evaluations have never come to the center of attention and stayed in periphery, to the point that the very existence of these MTs is not adequately known even among the academic body of translation studies in Iran.
The second factor is disinterested academics in the field. The fact that the general academic community in both translation studies and computational linguistics in Iran consistently shows a lack of enthusiasm towards

[DOI: 10.47176/kurmanj.6.4.1]

Machine translation research is undeniable. This is due to several reasons. Most notably perhaps is the common belief in Iranian academia that developers like Google and Microsoft are miles ahead of others and they will do the job for everyone! This makes any effort on developing MTs in Iran seem dead end and pointless even from the beginning. Notwithstanding the fact that specialized MTs, which are developed based on high-quality preprocessed parallel corpora, are always a much more reliable tool for translation. It is due to this fact that the mentioned Iranian MTs were developed in the first place and, by many standards, have made a positive impact.

The third factor contributing to the state of affair is the oblique utility potentials of MT systems in translation profession in Iran: Lack of interest in MT, in a systematic manner, has led Iranian MT research to a state in which even the potentials of utilizing MTs are not really being scientifically observed and valued. A series of systematic and scientific survey on the opportunities and threats resulted by this technology can be one of the most generative and profitable strategies.

Considering the above-cited factors, the second one, i.e. the academic disinterest, seems to play the most pivotal role. In this regard, a fresh perspective would be helpful. For instance, conducting futuristic research in the failed seems most valuable. These kinds of research need integrative and systematic planning rather than partial solitary studies.

Second, there is a fundamental void of suitable MT metrics for Persian-English language pair: the existing metrics are mostly developed based on close language pairs like English-French. Their assessment can be less relevant when dealing with different language pairs. In this respect, we need MT metrics that are applicable for Persian-English language pair.

Third, despite the efforts made on interdisciplinary cooperation, there is a substantial room for improvement in this area: many scholars advocate that translation studies itself is interdisciplinary. In the case of machine translation, there is no doubt that we have to consider linguistics, translation studies, sociocultural studies, statistics, data science, and computational linguistics, in order to make an integrated progress. Therefore, it is necessary that a department (or at least sub-department or committee) be in charge of facilitating inter-departmental and inter-organizational liaison.

Discussion And Conclusion

To sum up, the following suggestions can be put forward for a better progress in the field of machine translation evaluation research in Persian-English language pair:

Planning and developing integrated systematic research branches in the field by a focal office in charge of integrative research and inter-organizational liaison in the field of machine translation research and development can most likely increase productivity in further research. In managing a long-term project such as "MT development", strategic planning and strategic management techniques can be employed in order to devise a successful strategy for progressive development. In this integrated systematic research approach, an integrative attitude towards the entirety of the project would lead to a more productive progress. In this regard, a chart illustrating different branches of MT evaluation research areas can be devised and prompted by the academic body among both postgraduate students and established researchers in the field. In this manner, the research results can then be channeled to more productive meta-analyses and feedback for the MT developers.

Some of the main areas of research in Machine translation evaluation that can be put forward are:

•Micro linguistics: including syntax, morphology, semantics:

The focus is on evaluating the accuracy and adequacy of machine translated target text in the level of sentence and bellow. This requires analytic test suit designs in both synchronic and diachronic manners.

•Macro linguistics: including discourse and style, cross-cultural issues, and target text naturalness:

This is the hard kernel of the process. The linguistic level shall be the text and the semantic level is the massage as a whole. Because of the length of texts, developing AIs for the very purpose of Evaluating Macro level factors such as consistency is crucial.

•MTs utilization: focusing on post-editing and resource efficiency:

Research on post editing strengths, weaknesses, opportunities and treats, can help us to come up with productive post editing strategies. This means a great amount of study on identifying and categorizing various factors that affect the outcome of post editing process.

Implementing these measures may hopefully help and improve the utilization of machine translation, resulting in saving time, budget, and resource spent in both public and private sectors.

References

1. Koehn P. Statistical machine translation. Cambridge ; New York: Cambridge University Press; 2010. 433 p.

2. Hutchins WJ. Machine Translation: A Brief History. In: Concise History of the Language Sciences Internet.. Elsevier; 1995 cited 2022 Oct 2.. p. 431–45. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780080425801500660

3. Oladosu J, Esan A, Adeyanju I, Adegoke B, Olaniyan O, Omodunbi B. Approaches to Machine Translation: A Review. FUOYE J Eng Technol Internet.. 2016 Sep 30 cited 2023 Aug 9.;1(1). Available from:

[DOI: 10.47176/kurmanj.6.4.1]

https://journal.engineering.fuoye.edu.ng/index.php/engineer/article/view/26
4. Olive JP, Christianson C, McCary J, editors. Handbook of natural language processing and machine translation: DARPA global autonomous language exploitation. New York: Springer; 2011. 936 p.
5. Nunes Vieira L. Post-Editing of Machine Translation. In 2019. p. 319–35.
6. Sanchez-Torron M, Koehn P. Machine Translation Quality and Post-Editor Productivity. In: Conferences of the Association for Machine Translation in the Americas: MT Researchers' Track Internet.. Austin, TX, USA: The Association for Machine Translation in the Americas; 2016 cited 2023 Aug 27.. p. 16–26. Available from: https://aclanthology.org/2016.amta-researchers.2
7. Alvarez S, Oliver A, Badia T. Quantitative Analysis of Post-Editing Effort Indicators for NMT. In: Proceedings of the 22nd Annual Conference of the European Association for Machine Translation Internet.. Lisboa, Portugal: European Association for Machine Translation; 2020 cited 2023 Aug 26.. p. 411–20.Availablefrom:
https://aclanthology.org/2020.eamt-1.44
8. Koehn P. Neural Machine Translation Internet.. 1st ed. Cambridge University Press; 2020 cited 2023 Feb 13..Available from:
https://www.cambridge.org/core/product/identifier/9781108608480/type/book
9. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All you Need. In: Advances in Neural Information Processing Systems Internet.. Curran Associates, Inc.; 2017 cited 2023 Aug 10.. Available from: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
10. Marie B, Fujita A, Rubino R. Scientific Credibility of Machine Translation Research: A Meta-Evaluation of 769 Papers Internet.. arXiv; 2021 cited 2023 Dec 6.. Available from: http://arxiv.org/abs/2106.15195
11. Absalan SMJ. Comparing the output quality of Bing, Abadis and Farazin translation machines according to Dugast's model. Kerman Institute of Higher Education; 2020.
12. Ashrafi A. Investigation of Two Types of Machines Translations Google and Targman in Five Scientific Disciplines based on BLEU Model. 2022 Dec 4;9.

KURMANJ

Citation: Sadeghzadeh Yazdi F, Hashemi MR. Machine Translation Evaluation Research in Persian-English Language Pair: A Non-Statistical Meta-Analysis. KURMANJ, 2024; 6(4): 1-5.

https://doi.org/10.47176/kurmanj.6.4.1

[DOI: 10.47176/kurmanj.6.4.1]