

Innovations in Machine Translation: The Role of Machine Learning in Enhancing Linguistic Accuracy and Efficiency

Kaiwen Xin ^{a, #}, Bingchen Liu ^{b, #}, Lihao Fan ^c

College of Computer and Cyber Security, Fujian Normal University, Fuzhou, Fujian 350117, China

^a xin1311980670@gmail.com, ^b 121152021068@student.fjnu.edu.cn, ^c 121152021078@student.fjnu.edu.cn

[#] These authors contributed equally to this work

Abstract: This essay explores Instance Induction, Analogy Induction, and Machine Learning, with a specific focus on the application of analogy-based machine learning in machine translation. This includes full instance translation, case pattern translation, and analogical reasoning. The study examines the underlying principles, advantages, and potential limitations of these methods to provide a theoretical foundation for further optimizing machine translation (MT). Moreover, an in-depth examination of Machine Learning Theory, particularly through the paradigms of Analogy Induction and Instance Induction, is conducted to unearth latent patterns and features, which are pivotal for the technological evolution of this field. The efficacy of these methodologies in augmenting the performance of machine translation is critically analyzed and discussed.

Keywords: Machine translation; Application of machine learning technology; Artificial intelligence in Linguistics.

1. Introduction

With the increasing accessibility and user-friendliness of intelligent products, machine translation, a key element of artificial intelligence, exhibits substantial potential for expansion and evolution. The proliferation of machine translation software applications is a testament to this expanding field. Most of them are based on NLP rules, and they can be used to improve the efficiency of translation. In recent years, reinforcement machine learning has shown great promise in the field of machine translation. Thanks to machine learning, the machine has been able to improve its translation performance from a great deal of data. Also, machine translation software may be able to better understand linguistic context and semantic meaning, thus achieving more accurate and natural results. Therefore, it is very important to improve the intelligence and user experience of machine translation. MT will therefore assist in the application of MT and MT in the various fields and occupations. Through the combination of domain knowledge and machine learning, machine translation is more specialized and precise. Along with the technical progress and the popularity of machine translation, the improvement of the translation quality and the user experience has become more and more important. It will also bring the new chance and challenge to the smart development of MT and the application of industry.

The goal of machine learning is to improve the precision of complex logic by gathering historical data and taking the whole problem into account. As shown in Figure 1, this is a manifestation of machine learning and human thinking. By providing correct or incorrect examples of specific target concepts, the program analyzes and summarizes them to form a comprehensive conceptual system, thereby identifying and eliminating errors in new instances. Analogy based machine learning is a learning approach that is closer to peoples thinking and behavioral habits, that is, by comparing and identifying similar associations between different individuals. When new problems arise, the program will quickly review

and sort out historical data, sort and analyze similar cases previously handled, and find the most similar handling method to the current problem as a reference for new problems. The purpose of machine learning is to mimic the way people think and study. The application can study the right decision mode by applying the correct and wrong sample, and then make the correct choice. This type of computer learning, which takes people as its foundation, can easily handle complicated and changing circumstances. The Machine Learning Analogy aims at helping the computer to comprehend the messages that are present in history, and to discover the similarities and differences between them. Analogical study helps to comprehend and deal with complicated questions, mimic the way of thinking, enhance the AI of the program, and offer a more complete and deep reflection.

Computer Learning aims at simulating humans thoughts and learning, and then making the right decision-making on the basis of repetitive examples. This kind of machine learning, which imitates the way people think, makes it easier for them to handle complicated and changing situations, thus improving their ability to solve problems. The Machine Learning Analogy also makes it possible for the machine to comprehend the information in history, to discover the similar and the law, and offer the solving method. Analog Machine Learning can help us to comprehend and deal with complicated questions, to imitate peoples thoughts and to enhance the AI of programs. By analogy study, it is possible to gather and use large amounts of history information and information, especially when there are no explicit rules or rules. Analogue research enables the machine to be more complete and deep. So its very important for us to improve the systems intelligentization and resolve all kinds of actual problems. Simulating study, which is built on the basis of thinking and drawing historical data, enables the machine to be more agile and can deal with complicated problems easily.

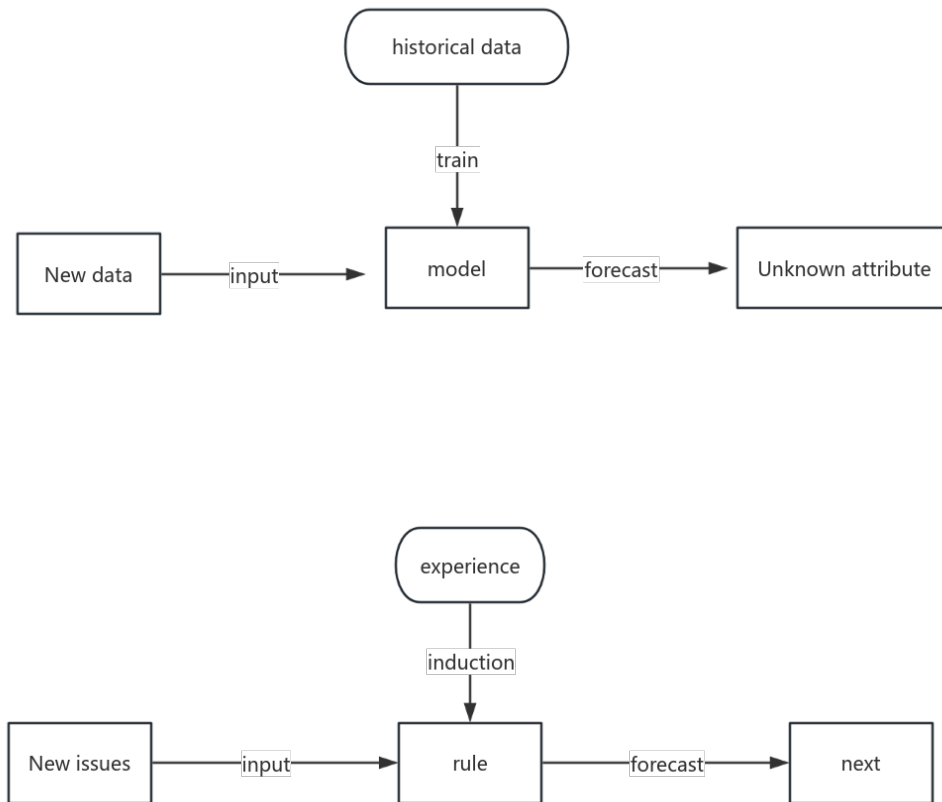


Figure 1. The mechanism between machine learning and human thinking is consistent

Table .1 Datasets available for software quality prediction

Dataset Name	Dataset source	Dataset Target
GitHub	GitHub MS	Defect/maintainability prediction
Cocomo v1 & 81	Jairus Hihn, JPL, NASA, Manager SQIP	Cost estimating
KC1	NASA	Defect prediction
JM1	NASA	Defect prediction
CM1	NASA	Defect prediction

2. Instance based machine learning

2.1. Technical ideas

Example based machine learning can be set according to single or cumulative methods. In a single behavior, examples of right and wrong will be presented at once; In terms of cumulative behavior, system programs can more conveniently accumulate instances through multiple assumptions and continuously revise and modify conceptual hypotheses to improve the accuracy of instances. The sample accumulation machine learning algorithm aims to simulate the human learning and thinking process, and technically approach human learning initiative as closely as possible. When machine learning begins, the processor will sort out various theoretical concepts it has mastered, directly expanding the connotation and extension of these concepts, and assisting the project in guiding the implementation of new cases. While understanding the new concept, it also achieves an understanding of the starting point of the new concept through the understanding of the new concept. From a practical perspective, the accuracy of cumulative behavior is generally

better than that of single behavior, and single behavior often reshapes concepts. In machine learning, both single behavior and cumulative behavior have their own advantages and disadvantages. A single action can quickly display correct and incorrect examples in a single action, but due to limited information, it can easily lead to a one-sided understanding of concepts; In contrast, cumulative behavior can comprehensively integrate historical data and cases by gradually accumulating instances multiple times, improving the completeness and accuracy of concepts. Therefore, cumulative machine learning is closer to the pattern of human learning, can better simulate the process of human thinking and learning, and improve the intelligence level of programs. In example based machine learning, both single and cumulative behaviors have their unique advantages and should be flexibly selected according to specific situations. Accumulative behavior can provide a more comprehensive understanding and processing of problems, which is more in line with human learning thinking patterns and is of great significance for improving the accuracy and intelligence level of machine learning.

Some incorrect examples are harmful to machine learning.

Example learning is a learning method based on inductive thinking, and strives to achieve a general description of the objective of the example. General explanation refers to the ability to derive new information from the results of an explanatory behavior based on its explanation [10]. When programming computer programs, it is necessary to use a certain standard description language to make the general description more formal. Any type of data input and description language can be generalized using a formal and

universal representation method to obtain a set of relationship descriptions with stronger constraints and more representativeness. The smallest element can be defined as the feature input value, and the largest element is called the typical description. Therefore, the usual approach is to first describe the characteristics of the input and then organize them through induction.

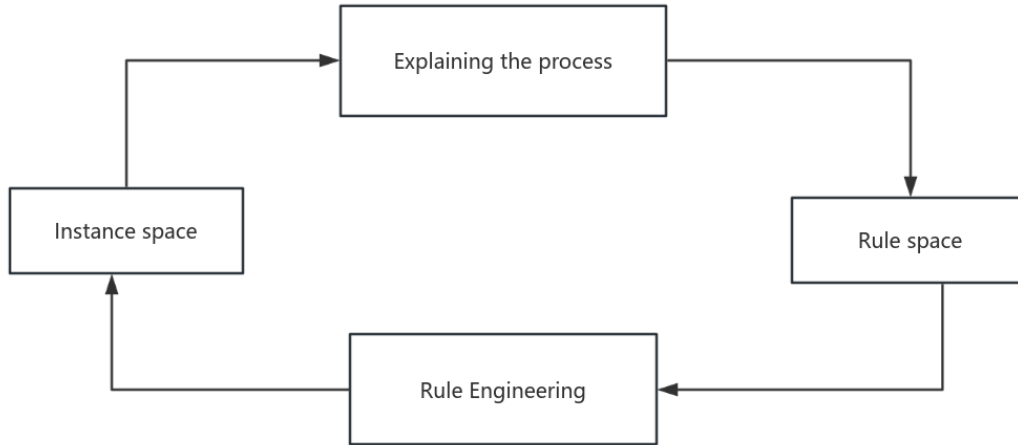


Figure 2. Schematic diagram of the system model for machine inductive learning

2.2. Case based problem description

In general, scripts, frameworks, hierarchies, rule generation, semantic structures, inference networks, and other methods can be used to describe the problem of events. In these three aspects, the study of dual calculus is more rigorous, and the generated syntax and semantics are also more refined. At present, a large number of artificial intelligence inductive learning tends to use predicate calculus, and theoretical research on inductive behavior has accumulated rich practical experience. Of course, there are many learning methods that limit how to make universal descriptions. For example, in a learning system, although it can be directly described, a description mechanism has not been established [11].

2.3. Case based machine learning execution rules

For general differential semi ordered descriptive bodies, specific ways from non general descriptions to general descriptions can be indicated. This requires us to conduct necessary induction and logical reasoning, and establish a unified execution criterion for all initial and intermediate variables, also known as the transfer criterion. Logically, it can be expressed as follows: when $S1 \rightarrow K$ appears in the general executable classification system, a more general classification system $S2 \rightarrow K2$ will be obtained. After comparing the descriptive text or quantization values of $S2$ and $S1$, if the formers descriptive content is completely mapped to the latters descriptor, it is called a selection rule. If the formers descriptive content has other descriptors outside the corresponding mapping, it is called a constructive rule. From this point, it can be seen that the emergence of constructive rules will inevitably lead to changes in the general descriptive space, while the rule of choice does not have this phenomenon. Lowering the conditional rule to the minimum is a lower level operation on general rules, which

achieves generalization by reducing the order of the composite rule. For example, "yellow bananas"[12].

After performing conditional rule subtraction, removing the conditional factor "yellow" in the classification will generate an all banana level without color. Promote it as: $\text{banana}(V) \rightarrow K$, and generalize it as: $\text{banana}(V) \rightarrow K$. The above simplified form is a selection rule, as there are no new explanatory elements appearing in the conclusion section of this explanation. In the process of generating constructive laws, "partial order polar elements" can be used to solve the problem. The core idea is to search for each node in the ordered chain structure and characterize it to improve its structure.

3. Analogy based machine learning

3.1. Technical ideas

In recent years, machine simulation learning has received increasing attention from scholars and has gradually become an important research direction in the field of artificial intelligence. Through the study of analogical learning methods, not only can we better grasp the learning forms of new knowledge, but also provide valuable reference for solving practical problems. At present, the ability that people need to possess in intelligent logic systems is mainly analogical reasoning. Through analogical reasoning, the similarities and contradictions between different things can be objectively and reasonably sorted out, and the corresponding attributes of other things can be derived from known attributes or interacting things. If the attribute of A is abc and the attribute of B is abcd, then the attribute of A is d. This indicates that machine learning methods based on analogical learning can effectively utilize existing system knowledge and, based on this, induce the laws and patterns of larger scale systems. The object elements that must be followed in analogical reasoning refer to the universality of the process of

an event occurring in a specific event and the relationships between its various elements.

3.2. Analogy based machine learning steps

The fundamental course of Computer Analogue Study is that when a new question comes up, the machine will be able to retrieve or connect to it. Then search for a similar question that is most closely related to the new one, because the more similarity there is, the better it will be. In this step, your machine will search for the most relevant related problem in the KB that will assist you in solving this new problem. In the last part, we build up the relationship between the two methods, namely, analogous and analogous [13]. The final stage is to examine the application of the new technology and knowledge to a new issue. When something goes wrong, it fixes it with the Analogy Learning Principle.

In specific operations, it is necessary to further strengthen the construction of the analogy system to provide more powerful guarantees. This means that machines need to continuously accumulate, organize, and optimize their knowledge base in order to more accurately find similar problems, correct errors in a timely manner, and continuously improve their analogical learning systems to adapt to various complex problem-solving scenarios [14]. The process of machine analogy learning is an iterative process, which continuously seeks similar problems, establishes corresponding relationships, adjusts and corrects errors, and improves the intelligence level of the machine, enabling it to better cope with the challenges of new problems. With the continuous development and improvement of machine learning technology, machine intelligence based on analogical learning will become more precise and efficient, providing more reliable support for solving practical problems.

4. Research on the Application of Machine Learning Technology in Machine Translation

This article mainly discusses the application of analogical learning in machine translation.

4.1. Applicable ideas

Analogy refers to the use of existing translations by predecessors to enhance the imitation of new things. In machine translation, knowledge source information containing a large number of translation instances is often used to connect with new translations, thereby achieving the correspondence between phrases and sentences in the translation. Although different approaches may result in differences in each step during this corresponding process, the language transformation of knowledge source information and target information is not affected in any way. The basic idea is to move from words to words, from phrases to phrases, from long sentences to long sentences, and from grammar trees to grammar trees. Simply put, this is a process of machine translation based on the idea of analogical learning. It can be expressed in this way: if a sentence S to be translated is given, then a sentence S that is most similar to the sample can be found in this sample library, and the translation of S is the translation of S. In the case of analogical learning, machine translation becomes a process of finding, corresponding, and reproducing. The closer the instance is found, the higher the translation accuracy of the new sentence

[15]. The machine must have a good grasp of the skills and characteristics of historical translation sentences in order to become a reference for new sentences. Analogy learning plays an important role in machine translation, enabling machines to better cope with new translation tasks by learning and utilizing historical translations. With the continuous accumulation of instance libraries and the improvement of algorithms, machine translation based on analogy learning will be more accurate and efficient, providing better support for cross language communication.

4.2. Format conversion in analogical learning

4.2.1. Complete sample conversion.

This translation form refers to the ability to find sentences that match the translated sentences from the established translation sample library after parsing the translated sentences. In this way, a complete example translation becomes a query behavior, and the included translation samples must be bilingual in both Chinese and English. In the process of establishing a database, the first step is for staff to combine some common sentences into bilingual text, and then input it into a module of the database or translation system. At the same time, it is necessary to ensure that the entire database can be continuously filled with translated text information, providing important reference samples for subsequent machine translation work [16]. In theory, if the sentence to be translated is composed of several commonly used phrases or phrases, it can improve the success rate and accuracy of retrieval. However, in reality, many natural languages are not composed of phrases and phrases, but of a large number of structural statements, which add grammatical and semantic differences between different statements. This information cannot be carried by a translation paradigm library or other computer system modules, which means that this complete example translation processing method does not have higher practicality. Of course, a complete example translation cannot meet the human demand for machine intelligence translation. However, in actual system development, we will continue to maintain this approach by storing massive, stable, and mature bilingual translation samples. This way, when needed, some key links in the translation can be effectively omitted, especially in automatic word segmentation, syntactic semantic analysis, translation conversion, etc. Therefore, the success rate of translating various types of sentences will also be significantly improved [17].

4.2.2. Syntactic conversion of example sentences

In the existing sample library, the number of bilingual papers included cannot reflect the total translation rate of all cases. Of course, from another perspective, replacing a complete example with a complete one can improve your success rate. The core idea is to structurally map certain words and patterns in longer sentences, so that the words and sentence patterns in the translation can be generated. In this case, the consistency of sentence structure will become the key to bilingual translation. In this format, the first thing to consider is whether the order of sentence transformation is consistent. In this case, the translation of sentence patterns needs to achieve both automatic segmentation and ensure grammatical correctness. On this basis, attention should also be paid to the different orders of syntactic transformations [18]. Compared to complete example translation, although example translation has high technical and practical value, it still cannot skip steps such as automatic word segmentation

and grammar semantic analysis.

4.2.3. Analogous translation

If neither of the above methods can solve this problem well, then the computer system is required to open an approximate translation logic mode. The translation concept of this format advocates for enhancing similarity matching between sentences, finding examples with the closest structure from the database, defining the similarity between the two, and selecting the best sentence by comparing the similarity between different sentences. In this case, the calculation of similarity becomes a key issue. The system can incorporate the calculation method of distance and implement the concept of punishment. The lower the score, the better. The system selects the formula with the lowest score for translation. It is worth pointing out that in order to reduce the operational complexity of analogical translation, efforts can be made to increase the penalty points, such as increasing or decreasing the elements of sentence structure expressions, and determining the penalty points by increasing or decreasing the elements. The specific deleted element can be determined in advance based on the existing pattern representation in the database. When adding specific elements, the penalty score will also be adjusted according to different parts of speech, and the specific operating mechanism needs to be improved [19].

5. Discussion

The essay systematically discusses the application of computer translation from the angle of case analysis and inference. Based on the above key techniques, we will systematically analyse the basic principles, strengths and limitations of this approach, and discuss its application in the field of IM. The aim of this study is to gain a deeper understanding of machine learning and to apply it to translation. The key point of this paper is to examine the theoretical basis of the model inference and the reasoning inference. This involves analysing its mechanisms, identifying their unique strengths, and revealing their potential limitations. Through this project, we hope to clarify the application of the algorithm in machine translation and promote its development. At the same time, this paper will also discuss the application of these principles to other theories of machine learning, which will be helpful to the development of translation theory. This project will reveal the intrinsic relationship between the theory of machine learning and its application. It will be helpful for us to understand the complicated relation of these two subjects and promote the development and innovation of MT. In this paper, we will make a deep study of the typical Machine Learning Methods, so as to find out the obvious rules and rules, and promote the development of MT. The results of this project not only have significant academic value, but also provide a basis for increasing the practicability and precision of machine translation. Finally, the results of this project will be applied to the actual application of MT and MT. By analyzing these problems deeply, this paper provides a new and revolutionary approach to machine translation. It is hoped that this project will provide a new way of thinking for MT and MT and MT and MT. The purpose of this project is to explore the complexity of examples and analogical reasoning, and to integrate them into machine translation research. Through careful analysis and deep theoretical research, we have contributed greatly to the study of machine translation, and

have laid a solid foundation for the future research.

References

- [1] Anning, Zheng Yongyan, 2020. Review of Language Awareness and Multilingualism [J]. Foreign Language Teaching and Research (3): 473-478.
- [2] Gao Lulu, Zhao Wen, 2020. Overview of Machine Translation Research [J] Chinese Foreign Languages (6): 97-103
- [3] Hou Qiang, Hou Ruili, 2021. Research on Neural Machine Translation - Insights and Prospects four[J] Journal of Foreign Languages (5): 54-59
- [4] Wang Huashu, Liu Shijie. 2021. Research on the Transformation of Translation Technology in the Era of Artificial Intelligence [J]. Foreign Language Mathematics (5): 87-92
- [5] Wang Huashu, Liu Shijie: 2022. Review of Translation Technology Research at Home and Abroad (2000-2021) [J]. Foreign Language Electronic Teaching (1): 81-88,9
- [6] Xu Jinfen, Pan Chenqian (2021). A Plan for Foreign Language Education with Chinese Characteristics under Multilingual Awareness [J7: Foreign Language Teaching (2): 49-54
- [7] Xu Hong, 2012. Research on Ethical Differences in Translation - Based on Literary Translation in China [MI. Shanghai: Shanghai Translation Publishing House,
- [8] You Jianwei, Li Mingming, Shi Lan [2021]. Exploration of English Teacher Beliefs in the Context of Translation Technology [J]. Shanghai Translation (3): 44-49.2021.
- [9] Zheng Yongyan, Liu Weijia The Composition of Multilingual Motivation and Cross linguistic Differences of Chinese Learners [J]. Foreign Language and Foreign Language Teaching (6): 45-57148
- [10] MAHMOOD Y, KAMA N, AZMI A, et al. Software effort estimation accuracy prediction of machine learning techniques: a systematic performance evaluation [J] . Software: Practice and experience, 2022 (1): 39-65.
- [11] COWLESSUR S, PATTNAIK S, PATTANAYAK B. A review of machine learning techniques for software quality prediction[J] . Advanced computing and Intelligent Engineering, 2022: 537-549.
- [12] NORMAN E Fenton, SHARI Lawrence Pfleeger. Software metrics: a rigorous & practical approach [M] . Boston: PWS Publishing Company, 1997.
- [13] SINHA H, BEHERA RK. Supervised machine learning approach to predict qualitative software product [J] . Evolutionary Intelligence, 2021 (14): 741-758.
- [14] RANATHUNGA S, LEE E S A, PRIFTI SKENDULI M, et al. Neural machine translation for low-resource languages: A survey[J]. Association for Computing Machinery Computing Surveys, 2023, 55(11): 1-37.
- [15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. California :NIPS, 2017: 6000-6010.
- [16] BOHAN LI, YUTAI HOU, WANXIANG CHE, et al. Data augmentation approaches in natural language processing: A survey[J]. Ai Open, 2022, 3: 71-90.
- [17] JACOB D, MING C, KENTON L, et al. BERT: pre-training of deep bidirectional transformers for language understanding. [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association Computational Linguistics. Stroudsburg: ACL, 2019: 435-470.
- [18] CHEN G, CHEN Y, WANG Y, et al. Lexical-constraint-aware neural machine translation via data augmentation. [C]//

- Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, Yokohama: IJCAI 2021:3587-3593.
- [19] SAUNDERS D, STALHBERG F, et al. Multi-representation ensembles and delayed SGD updates improve syntax-based NMT[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics . Australia : ACL.2018: 319-325..
- [20] POST M. A call for clarity in reporting BLEU scores[J]. arXiv Preprint, arXiv: 1804.08771, 2018.