MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING – A REVIEW

Tejenderkaur H. Sandhu

ME Scholar, Department of Computer Science & Engineering

Sipna College of Engineering & Technology, SGBAU

Amravati, India

Abstract: Machine Learning is the part or subset of Artificial Intelligence (AI) that uses computerized reasoning and studies and estimates how to perform the task using the previous inputs and experiences to solve the task without being unequivocally modified. Natural Language Processing is a subset branch of Artificial Intelligence (AI) that mainly tries to build the system that can understand human language and generate and analyze output. This paper describes what these two concepts consists of and how they acts intelligently for performing tasks.

*Keywords***:** Machine Learning, Natural language processing, Artificial Intelligence, Algorithms.

1. INTRODUCTION

Artificial Intelligence is a knowledge that is utilized to recreate the innovations that allow machine to do work in a smart way. AI is a branch of the computer that incorporates thinking and critical thinking exercises, observation and learning, normal dialect preparing information portrayal and use of objects.It is the learning of how to make a framework which can think, act and erform precisely or superior to what an individual human can act or respond. Artificial Intelligence has advanced out of from areas such as linguistics, philosophy, mathematics and psychology. [1][2][10]

Machine Learning is a part of Artificial Intelligence which aims to build the systems that can work with past performances or previous examples. As it is the subset of AI it has to work smartly for the better output. Machine Learning contains many problems in AI. It uses these problems to solve other one's in AI. Aritificial Intelligence requires such set of learning beacuse Machine Learning tends tolearn itself which makes it an intelligent behavior machine; intelligent machine requires knowledge and learning acts a better source of knowledge.[1]

The Machine Learning field advanced from the wide field of Artificial Intelligence. The most important aspect is how to make the system to learn itself without human help. It improves the system with the help of data it has already, knowledge it has before and the one which it develops after experiences and interactions.

Mahine Learning is development of algorithms and techniques instead of programming for performing functions. It can be used for Natural Language Processing, making interfaces, decision trees, statics, knowledge representation, AI problems, planning and controlling, computer system and vision and also for databases. All Machine Learning comes under AI but not all AI are Machine Learning.[4]
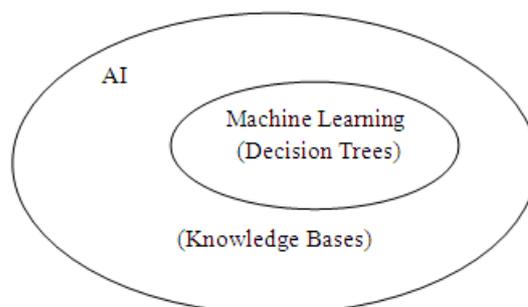
Figure 1: Machine Learning

Natural Language Processing is the strategy used for dealing with interactions between computer and human. It uses theories and technologies for generating the relevant output. NLP is basically a way that human can use for achieving its task from computer without actually programmimg it; instead making it sensible so that it can understand the language as a form of input to the system and later analyzing it and giving desired output.[1][9]

Natural Language Processing uses Machine Learning algorithms to solve the problem. It builds up the system that understands the language using learning experiences. There are some important levels of NLP:

1) Phonology: It is the study of sorting sounds through system.
2) Morphology: It is the development of words through primitive meaningful units.
3) Lexical: It is the studying and identifying structure of words in paragraph.
4) Syntactic: It is the analysis of words in a sentence in order to find out the relationship between them by arranging it in a manner.
5) Semantic: It checks whether the presented word has a correct meaning or not.
6) Discourse: It is responsible for analyzing the meaning of current, previous as well next sentence.

7) Pragmatic: It tries to understand the real meaning of the sentence that is said. It is responsible for reinterpreting.[5]
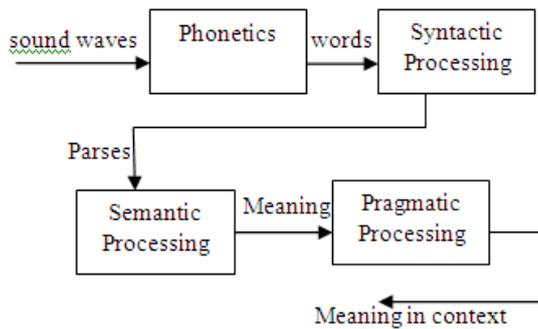
Figure 2: Natural Language Processing

2. MACHINE LEARNING TECHNIQUES

Supervised Learning: In Supervised Learning, the information variable and yield variable are given. The yield of the given information is relatively right in light of the fact that here is an educator accessible to teach the way toward taking care of the issue as per the given info.

The information variable X and yield variable Y exist in the supervised learning. The strategy for managed learning is quick and precise. It has the best possible readiness of preparing, approval and informational indexes. It is the way toward learning and calculation that is utilized to make a forecast on the given preparing informational indexes and it is adjusted by the educator. This procedure is halted when it accomplishes the normal or worthy yield. Regulated Learning is the most effective method to preparing choice tree and neural systems.

Unsupervised Learning: In Unsupervised Learning, there is no educator accessible and it has just info information variable. It is the way toward taking in the information with no direction. Unsupervised learning is a self - assessment calculation reviews process by a ton of ways, it has redressed and inaccurate yield accomplished from the info assessment. Unsupervised learning issue isolated into two issues initially is Clustering and second are Association-means and Apriority Algorithm is the mainstream calculation of Unsupervised Learning.[1][3]

3. ALGORITHMS

K- Nearset Neighbor: The least difficult strategy is the k-Nearest Neighbor classifier. Here the k purposes of the preparation information nearest to the test point are found, and a name is given to the test point by a greater part vote between the k focuses. This technique is very natural and accomplishes – given its effortlessness – strikingly low arrangement blunders, yet it is computationally costly and requires a substantial memory to store the preparation information.

Decision Trees: Decision trees are straightforward yet successful characterization plans for little datasets. The computational unpredictability scales ominously with the quantity of measurements of the information. Vast datasets tend to bring about convoluted trees, which thus require a substantial memory for capacity.[6]

SVM: SVM is another technique for the grouping of both direct and non-straight information. SVM are regulated learning models and it related with learning calculations that investigations information and perceives designs. The fundamental SVM takes an arrangement of information, for each given information, which has two conceivable class frames the yield making it a non-probabilistic parallel straight classifier.[7]

4. RECENT ADVANCEMENT

The most recent ten years of the thousand years have seen gigantic development in the field which can be credited to an expanded accessibility of a lot of electronic content, rapid PCs with bigger memory and the approach of the Internet. The gigantic changes in machine innovation have likewise enabled NLP to be performed with a genuine building soul. The present propelled processing assets permit the running of intense frameworks and directing of stunningly vast examinations in a brief term of time.

In the region of dialect handling, advance was made in the field of language structure with the improvement of successful punctuation portrayal means and procedures like diagram parsing. Accessibility of broad reasonable instruments likewise added to the assembling of numerous frameworks or interface subsystems for exploratory and improvement purposes. Measurable methodologies prevailing with regards to managing numerous nonexclusive issues in computational etymology, for example, grammatical feature distinguishing proof, word sense disambiguation, and so on., and have turned out to be standard all through NLP. NLP specialists are presently creating cutting edge NLP frameworks that arrangement sensibly well with general content and record for a decent segment of the inconstancy and vagueness of dialect. Discourse understanding frameworks with dialect preparing abilities would now be able to be thought of as the following level in this field.

Word2vec was another achievement in the field of NLP which has helped in surfacing semantic connections by unsupervised preparing of a lot of content. This calculation contains two unmistakable models (skip-gram and CBOW). These models have two diverse preparing techniques (with/without negative testing). Skip-gram demonstrate essentially predicts neighboring words around a word. Conversely, the CBOW demonstrate predicts the present word, given the neighboring word. Utilizing basic vector contrast, word2vec showed that it is conceivable to improve word portrayals if a model's multifaceted nature is exchanged for productivity. An adjustment of word2vec is doc2vec, an unsupervised calculation which creates vectors for words. These vectors are then utilized for discovering closeness between sentences. Doc2vec calculation is a perfect decision when the info corpus has been edit and has next to zero spelling botches.

Another prominent progression has been the approach of rough string coordinating, otherwise called 'fluffy string seeking'. In this calculation the precision of a match is dictated by the quantity of crude tasks (otherwise called alter remove) required to change over a string into a correct match. This calculation is broadly utilized as a part of spelling amendment projects to locate the best counterpart for a content not found in word reference.[8]

5. CONCLUSION

Machine learning approaches connected in methodical surveys of complex research fields, for example, quality change may aid the title and dynamic consideration screening process. Machine learning approaches are specifically noteworthy considering consistently expanding seek yields and openness of the current proof is a specific test of the exploration field quality change. Expanded analyst assention seemed, by all accounts, to be related with enhanced prescient execution.

Human level or comprehensible normal dialect handling is an AI-finish issue. It is proportionate to taking care of the focal computerized reasoning issue and influencing PCs as clever as individuals with the goal that they to can tackle issues like people and think like people and in addition performs exercises that people can't perform and making it more productive than people.

NLP's future is firmly connected to the development of Artificial knowledge. As characteristic dialect comprehension or decipherability enhances, PCs or machines or gadgets will have the capacity to gain from the data on the web and apply what they realized in reality. Joined with common dialect age, PCs will turn out to be increasingly equipped for accepting and giving helpful and ingenious data or information.

6. REFERENCES

[1] Archana Singh, Raj Shree "Recognition of Natural Language Processing to Manage Digital Electronic Applications", Vol 8, Issue 5, 2017, IJARCS, pp. 1641-1643.

[2] Nikita A. Kangude, Sanil B. Raut, "Introduction to Artificial Intelligence", Vol 2, Issue 3, 2012, IJCSET, pp. 958-962.

[3] Anish Talwar, Yogesh Kumar, "Machine Learning: An artificial intelligence methodology", Vol 2, Issue 12, Dec 2013, IJECS, pp. 3400-3404.

[4] Kajaree Das, Rabi N. Behera "Survey on machine Learning: Concept, Algorithms and Applications", Vol 5, Issue 2, 2017, IJIRCCE, pp. 1301-1309.

[5] Unnarti Dhavre, Prof. Umesh Kulkarni, " Natural Language Processing using Artificial Intelligence", Vol 4, Issue 2, 2015, IJETTCS, pp. 203-205.

[6] B. R. Patel, K. K. Rana "Survey on Decision Tree Algorithm For Classification", Vol 2, Issue 1, 2014, IJEDR, pp. 1-5.

[7] C. Cortes and V.N. Vapnik "Support vector networks. Machine Learning", Vol 20, Issue 3, 1995, pp. 273-297.

[8] Bhargavi Goel "Developments in The Field of Natural Language Processing", Vol 8, Issue 3, 2017, IJARCS, pp. 23-28.

[9] Dr. Khaled M. Alhawiti "Natural Language Processing and its Use in Education", Vol 5, Issue 12, 2014, IJACSA, pp. 72-76.

[10] D.B.Fogel, " Review of Computational Intelligence: Imitating life", Vol 83, Issue 11, 1995, IEEE, pp. 1588-1592.