ATHENA RESEARCH CONFERENCE
NEVER STOP QUESTIONING

A novel approach to Computational Linguistics- A comparison of the processes of Machine Learning and Language acquisition- and its implications.

Sohum Bajaj
The American Embassy School

ABSTRACT

The intersection between linguistics and machine learning is a field that has only recently gained popularity. However, due to its stark disparity of quantitative and qualitative methods, not much insight has been obtained through the intersection of the 2 fields of study. This study, aims to address that problem through a comparative approach of the process of first language acquisition, having its implications in the field of psycholinguistics, with the aim for 21st century parents to follow effective practices for their children to achieve fluency of their respective first language. The results demonstrated 5 thematic comparisons of the implications of machine learning in language acquisition. Through the findings of the study, machine learning and its process helps one bolster one's understanding of how to systematically and effectively "train" or teach a child their first language with the goal of achieving fluency, maximizing the accuracy, rate, and prosody of language.

INTRODUCTION

The aim of this research study is to promote the interest in the study of computational linguistics through a comparative analysis of the process of machine learning vs the process of first language, language acquisition in humans: the process by which humans learn the first language they know, usually from 0-4 years. Through studying a relatively complex domain like Machine Learning, whose many processes have their roots in human psychology and neurobiology, humans can learn from its implications presented through a comparative analysis. The research paper uses past research on the two "distinct" fields as a basis for a comparative analysis on how knowledge of one can augment and improve our knowledge of the other field. In the end, steps are provided by which Machine learning and its process can help a child acquire language and gain fluency with varied use of language. The results demonstrate how insights from Machine Learning can be used in the field of language acquisition in the aim to achieve a higher fluency.

METHODS AND MATERIALS

The research is a comparative analysis that uses past research as evidence to draw on connections between language acquisition and machine learning.The research was executed in 3 subparts- background research, question-specific research and then the comparative analysis. The background research consisted of papers purely on machine learning and language acquisition individually. Furthermore, research was done on the field of computational linguistics, where there were a few studies that involved the interplay between Machine learning and language acquisition() . After exhausting all the possible research papers, the comparative analysis was conducted, wherein the information from the question-specific research and evidence research was utilized to draw connections between Language acquisition and Machine learning.

RESULTS

1- The first similarity between first language language acquisition(FLLA) and Machine learning is what drives the transition to a next stage of data. Moving from stages from Pre-Speech all the way to Multiword stages, data is what drives the transition between each stage

2- Refining Data input- In machine learning, cleaning data aims to improve the accuracy score of the model by removing duplicates, normalization,and identifying points with missing values.In the same way, a parent aims to obtain a more notable, profound response by refining their questions which in FLLA signals a transition to stage 2.

3- Reinforcement Learning- An ML model's performance reflects its knowledge or the data it was trained on. Moreover, in reinforcement algorithms, an AI learn through rewards and penalizations in pursuit of obtaining an action with the highest point reward. This is a reinforcement process, used in many practical scenarios to obtain a high point reward, therefore minimizing the punishment. A child also uses reinforcement learning, to learn actions and words. This is done usually through learning through stimuli(experiences) actions, and the occurrence of favorable or beneficial events, called rewards, or unpleasant/ unfavorable events called punishments.

4-set barriers of stages- FLLA, the higher the rate of input, or the more data gathered, the faster the child will transition through the stages, ultimately achieving fluency. Various studies in psychology have demonstrated a positive correlation between exposure, and the rate of language acquisition.ML-higher input means higher accuracy

5- Machine learning uses an 80-20 split to reduce overfitting- By exposing children to various sources of language, parents can help the rate at which fluency can be attained. Moreover, to prevent overfitting, by exposing children to a variety of sources, parents can achieve an optimal fit(as shown in figure 1) with varied vocabulary usage.

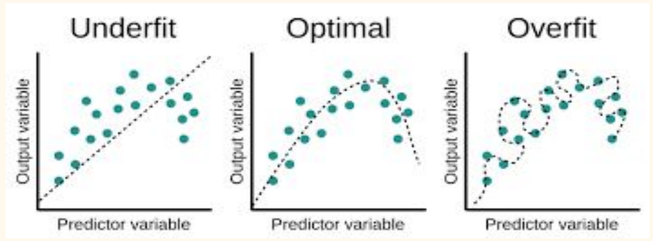

Figure 1.

DISCUSSION/Limitaion

The results reported herein should be considered in the light of some limitations. Primarily, because the field of computational linguistics is not very popular, and has very few papers on it. Since the relationship between ML and FLLA is ever more specific, this paper is based on limited research from a restricted amount of sources. The field of computational linguistics, however, is an evolving field, and as new information is gathered, more novel insights can be made about the two fields. Therefore, this research can not be used as a definitive data source, and will have to be supplemented by other research papers, confirming the findings.

Further research can be conducted on finding more connections between the two fields, possibly empirically which would increase the validity of the findings. Moreover, it can focus on one aspect of the study in detail as this study focuses on 5 different thematic comparison.

CONCLUSIONS

The paper demonstrates Machine learning and first language language acquisition are not two disparate fields of study and are more interconnected than we previously believed. Parents, aiming for their child to gain early fluency with a high range in vocabulary can therefore use the insights of the study to do the same- for example, parents in step 5 can aim to provide both "quality" and "quantity" as that indicates a higher fluency and large vocabulary.

REFERENCES

1. Al Zoubi, Samer. (2018). The Impact of Exposure to English Language on Language Acquisition.
2. Langley, P., & Carbonell, J. G. (1986). Language Acquisition and Machine Learning.
3. Hart, B., & Risley, T. R. (1995). Meaningful differences in the everyday experience of young American children. Paul H Brookes Publishing.
4. Hutauruk, B. S. (2015). Children first language acquisition at age 1-3 years old in balata. IOSR Journal of Humanities and Social Science (IOSR-JHSS), 20(8), 51-57.
5. Mayo, M. M. (2018, May). The 7 Steps of Machine Learning. KDnuggets. Retrieved October 7, 2021, from https://www.kdnuggets.com/2018/05/general-approaches-machine-learning-process.html
6. Sachowski, J. (2019). Implementing Digital Forensic Readiness: From Reactive to Proactive Process, Second Edition. CRC Press.