INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

IJARIIT

**ISSN: 2454-132X**
**Impact factor: 4.295**
***(Volume 4, Issue 5)***
Available online at: www.ijariit.com

# Artificial intelligence in word prediction systems

*N. Ajithesh*
*ajithesh.n@gmail.com*
*SRM Institute of Science and Technology, NCR Campus, Delhi*

## ABSTRACT

***Artificial Intelligence is the process of performing tasks that generally requires human intelligence by artificially emulating it in a computational background. Implementation of this concept in predicting the next suitable word efficiently such that it does not deviate from its original meaning. It is observed that most word prediction algorithms involve comparison of a word with a dictionary or a collection of words that shows resemblance and follow a fixed linear path by recognizing patterns. In most cases, this linear analysis solves the problem of maintaining the sentence structure and its meaning. However, in complex cases, it is found to be inaccurate as minor details of a language are ignored. This can be corrected with the help of machine learning along with pattern recognition by analyzing the frequency of the words, addition of new words into a local dictionary, identifying the difference between official and unofficial text sentences or words. It becomes more complex as the neural network keeps increasing but it improves and provides accurate results over time and can be improved with comparison using a cloud-based dictionary. The concept being implemented is universal and is useful in any kind of language and provides a restriction free communication and understanding.***

***Keywords*****—** *Artificial intelligence, Word prediction systems, Neural networks, Machine learning*

## 1. INTRODUCTION

Word prediction is an interesting concept that requires a computer to emulate the thinking aspect of the human brain to predict the next most suitable word while typing a sentence. It reduces human effort and improves the sentence itself grammatically, syntactically and makes it look more professional and concise. The scope is set to all the words that belong to the language "English" and the dictionary being "Oxford English Dictionary". The reference dataset consists of well over 171,476 words in current use, 47,156 obsolete words and 9,500 derivative words. The words can be formal as well as informal depending on the context it's being used. It is observed that informal words prevail over the other due to its frequent occurrence [1].

### 1.1 Informal words

SMS language is the most commonly used slang that has its usage in mobile phone text messaging or other internet-based communication [2]. It abbreviates formal words into more suitable short forms that are sometimes considered as homophones. They are unofficial words and are only often observed in text or Internet-based communication. After theoretical observation, SMS slang can be any of the following:

- Initialisations
- Reduction
- Pragmatics
- Paralinguistics and prosodic features
- Variations in spelling
- Punctuation, or lack thereof

### 1.2 Formal words

Words that can be found in any English dictionary also has its own definition [3]. All words belong to a set of classes known as "Word classes" or parts of speech based on the part they play in a sentence. [4] After theoretical observation, Word classes are as follows:

- Noun
- Verb
- Adjective
- Adverb
- Pronoun
- Proposition
- Conjunction
- Interjection

### 1.3 Sentence formation

For a sentence to be formed it must at least have a subject and the main verb state or declare a complete thought. [5] A subject is the noun that is doing the main verb. The main verb is the verb that the subject is doing. A phrase or clause is part of a sentence two clauses are joined together with conjunction or punctuations. It is hence observed that word classes have their own positions in a sentence to make it meaningful according to rules as follows:

- The simplest sentence consists only of a **noun**, a naming word, and a **verb** or action word.
- **Adjectives** describe nouns. Adjectives usually come before the noun
- **Adverbs** describe verbs. The adverb should always follow the verb.
- **Verbs**, or action words, are expressed in tenses; past, present or future.

## 2. METHODOLOGY

Machine Learning is the key to AI's Intelligence. There are many types of Machine Learning Algorithms. [6] In this case, Rule-based Machine learning is found to be promising since the base analysis mechanism is Pattern Recognition. It is a general term for any machine learning method that identifies, learns, or evolves `rules to store, manipulate or apply, knowledge (i.e.) data.

### 2.1 Local dictionary (5 step approach)

The dictionary is present locally and is accessible offline. In this phase, the written word is checked syntactically with the help of pre-existing reference sentences and is auto-corrected if there exist any verbal or logical errors. Once the error is recorded the following word is then analyzed for processing.

**Step 1: Frequency of Occurrence**

To predict the correct word that is suitable for the sentence apart from what follows the sentence consists of many combinations of words. This can be solved through identification of word frequency obtained from previous input data. It narrows down the number of words to a minimum from which it is easier to identify the next word thereby creating a frequency list for words.

**Step 2: Probability of the Data**

The frequency of a word is determined through computation of the probability of the words. Since ML depends upon mathematical properties for analysis and decision tree. Initially, all the words in the dictionary are given numerical value "1" as their probability index. Considering that computation takes whole number inputs. So, if there exists a word which has the least probability, it is hence obvious it is closer to zero after calculation of its probability. Arranging the probability values in decreasing order gives the words from most recently used to least used.

**Step 3: Slang and Inclusion**

There are words that generally are not found in Dictionary but are used as an abbreviation for a larger meaning of words (slangs). These words are most frequently used and have greater probability value than average dictionary words. These words need to be included in the reference dictionary to create the frequency list with their own probability values.

**Step 4: Optimization and Updating**

Frequency list needs to be properly managed and requires constant computation of probability values for its list of words. This improves the feasibility and accessibility of different word usage.

Implementation of different methods to reduce space and Increase speed for such prediction is an evolving process which directly affects the ability of ML in word prediction systems.

**Step 5: Matching of Words**

Identification of relationship between dictionary words and words in the frequency list can be tricky since the frequency list are not arranged in alphabetical order. They are arranged in accordance with their computed probability values. Each of the Input words is scanned to an array letter by letter. Each letter in the array is then used to sort the words from the dictionary. Finally, the sorted words from the dictionary are referred towards the frequency list to identify the most possible word. Once identified, the top three words are taken into consideration in determining the correct word for the sentence that gives it meaning both syntactically and grammatically. The thus identified word is given for prediction.

### 2.2 Zipf's Law

The above method for calculation of probability values for words belonging to any language is given by George Kingsley Zipf [7].

It is an empirical law which states that given a large sample of words used, the frequency of any word is inversely proportional to its rank in the frequency table. So, the word number N has a frequency of 1/N. Thus, the most frequent word will occur about twice as often as the second most frequent word, three times as often as the third most frequent word.

### 2.3 Observations (Local)

The above data is more than enough to give a biased prediction that is 58% accurate without using machine learning or neural networks. A graph is obtained which shows an accuracy variation averaged at 50-58%. This can be further enhanced without using neural networks but using a local machine learning algorithm.

Analysis of different words and their Accuracy rate in sentences with Probablistic values (without Local Dictionary Algorithm).

| Words | Probability | Accuracy |
|---|---|---|
| Adamant | 0.43 | 32.3% |
| Mischief | 0.11 | 26.32% |
| What | 0.88 | 56.1% |
| Return | 0.74 | 43.6% |
| Hi | 0.91 | 58.23% |

The above analysis is with reference to number of occurances in 100 different sentences, each word's probabilistic values are calculated.

Program used : Basic sentence prediction in Microsoft word (Predictive Text).

**Fig. 1: Image generated via Predictive text algorithm as mentioned above**

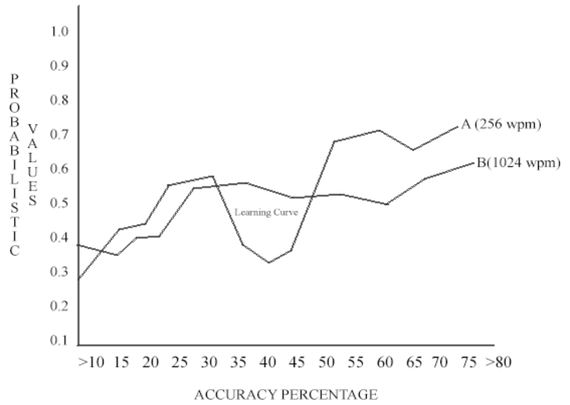


**Fig. 2: Probability vs. Accuracy w/ Learning Curve**

### 2.4 Rule-based machine learning (with respect to probability values)

A Decision tree [8] is constructed with respect to probability values. Three base values are used for constructing the tree.

1. Greater than 0.5
2. Lesser than 0.5
3. Equal to 0.5

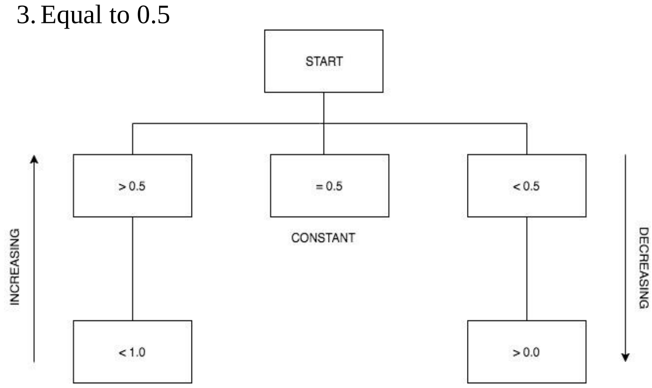


**Fig. 3: Algorithm for decision table localized**

**2.4.1 Interpretation of the table:** Three probability-based word lists are generated with reference to the table as High Frequency, Neutral and Low Frequency. The words that have a probability greater than 0.5 and less than 1.0 (High Frequency) are arranged in increasing order. These words are processed in such order and if the word chosen is not present in the list, it switches to the Neutral list and then to Low-Frequency list until it finds the chosen word in the list. Similarly, the words that have probability lesser than 0.5 and greater than 0.0 (Low) are arranged in decreasing order. These words are processed in such order only when the word chosen is present in the list. Once the chosen word is found and processed, the probability value of the word and its synonyms present elsewhere are increased "0.X" amount and the values other words are decreased by the same "0.X" amount. In an exceptional case where the word is not present in any of the lists, the word is given a Neutral value and is added to the table. If the value of any word is changed, the entire table is reconstructed with those new values. This is one complete cycle or epoch. This cycle repeats whenever local prediction is enabled.

**2.5 Cloud Dictionary (Neural Net)**
The dictionary is present in the cloud which is common to all the users available globally. This dictionary consists of words and sentences both new and approved by Oxford. In comparison to those words, it creates a neural network to analyze and process the information.

**2.5.1 Estimation of probability using neural language models:** It is observed that as the number of unique words increases the possible sequences also increases exponentially. [9] Data sparsity occurs as there might be exponentially larger sequences of data and that such amount of statistics is required to estimate the probability. Neural Network represents words as nonlinear combinations of weights in neural net in order to overcome such sparsity issues. Thus, the resulting Neural network might be either of feedforward or recurrent. Neural network language models are trained as probabilistic classifiers that predict probability distribution.

Initial neural step for the contextual approach:

$$P(w_t|context)\forall t \in V$$

[11] In order to achieve such probabilistic classification neural net training algorithms such as stochastic gradient descent with backpropagation is used with feature vectors representing the previous 'k' words in case the context is a fixed-size of previous words or the neural network model can be reversed to learn context from the word which maximizes the log probability.

Neural step for 'k' words with 't' possibilities.

$$P(w_t|W_{t-k}, \dots \dots w_{t-1})$$

Neural step for 'k' words with 't' possibilities having 't+1' to 't+k' exceptions: [12]

$$P(w_t|W_{t-k}, \dots w_{t-1}, w_{t+1}, \dots, w_{t+k})$$

Neural step for 't' possibilities with 'j' recurrences. [13]

$$\sum_{-k\le j-1, j\le k} logP(w_{t+j}|\, w_t)$$

**2.6. Observations (Cloud)**
The above approach was implemented with a simpler constraint among five different users have set up a common global dictionary on a server. It is identified that the probability index for the same list of words was much lower than the local observation. This further concludes the presence of larger database and wide variations in usage of same words among different users. However, the accuracy rate for the same is found to be 80-81% and the variations in word usage averaged 64%. A graph is obtained that shows variations with respect to accuracy for words used by varied users.

| | | |
|---|---|---|
| get_result(adamant) | 0.09 | 76.40% |
| get_result(what) | 0.63 | 79.32% |
| get_result(mischief) | 0.02 | 89.11% |
| get_result(return) | 0.44 | 88.14% |
| get_result(hi) | 0.92 | 97.39% |

**Fig. 4: [Left to Right] Server: mongodb function, probabilistic values, and accuracy percentage**

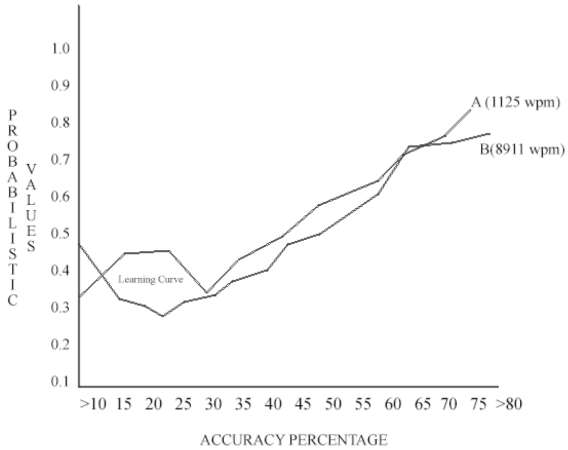


**Fig. 5: Probability vs. accuracy w/ learning curve**

**2.7 Neural network model**
Sentence analysis by breaking down words from the raw sentence and identifying its word class and its placement for making the sentence with reference to a similar sentence in the data gathered and estimating its accuracy rate. The inputs will be the words from the raw sentence with unique weights. The reference output will be the approved sentences from Oxford.

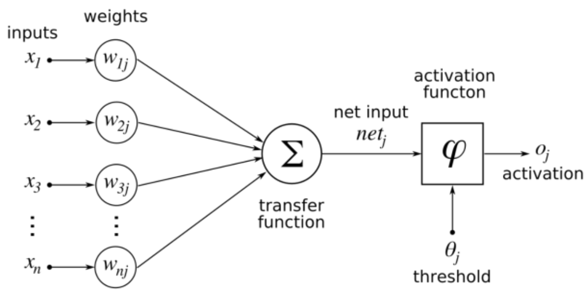


**Fig. 6: Basic weighted transfer function diagram [14]**

Word analysis by identifying its placement in a sentence including its synonyms and its usage with reference to word data gathered and estimating its accuracy rate. [15] The inputs will be the words collected from the user's High, Neutral and Low-Frequency lists with unique weights. The reference output will be the approved words from Oxford.

**2.7.1 Interpretation of the model:** The locally processed information and the word lists are sent to the cloud for neural network processing and machine learning. The local word list already consists of probability values which can be used to attain a head start in neural network processing by giving priority to the highest probability values. The entire sentence or paragraph in which the next word to be predicted is considered and is sent to the cloud. The Neural network model which does the breaking and understanding of the sentence increases the probability of the next suitable words thereby optimizing the table and removing inconsistencies. The output word with highest accuracy rate is sent to the user for prediction. The sentence or paragraph is added to the cloud's database as a reference data. The Neural Network model is constantly optimized for better machine learning and the data is completely stored in the cloud reducing storage restrictions and improving efficiency.

## 3. PERFORMANCE COMPARISON (LOCAL VS CLOUD)

A table was generated to analyze the difference in various parameters between local – 5 step approach and cloud – neural network approach.

**Table 1: Values are approximated to nearest decimal values except for recurring digits**

| Factors | LOCAL | CLOUD |
|---|---|---|
| Dictionary | Oxford | Oxford |
| Editable | Yes | No |
| Presence | Disk | Server |
| Slangs | Yes | No |
| Localization | Yes | No |
| Probability Rate | High | Low |
| Accuracy | Low | High |
| Size | Fixed | N/A |
| Words Per Minute | 1024 | 8911 |
| Difference in Accuracy | 31.33% | >10% |
| Algorithm | Rule-Based | Neural Network |
| Efficiency | High -> Low | Low -> High |
| Optimization | Low | High |
| Load on Device | High | Low |

## 4. CONCLUSION

Artificial Intelligence with the help of Machine learning algorithms is found to be implemented in various output derived fields where statistics and predictions play a major role. However, the way of implementation of such algorithms is directly proportional to the efficiency of the system thus developed. Word prediction systems have played a major role in any computational devices but the presence of algorithms and database required for such high-end computation for achieving accurate results is within the devices itself thereby decreasing the efficiency and load on the systems. This was the only way of implementation due to insufficient speeds between client and server. In recent years where data transfer speeds are greater than 100 Mbps, it is easier to achieve such results without having the database on the device itself. This improves the entire system making it more secure and reduces the risk of hardware degradation.

## 5. FUTURE SCOPE

Languages and meanings differ between various regions of the world. Interpretation and variations of the same language are a commonly observed the phenomenon. The above approach and methodology can be further enhanced based on the region in use. The language "English" is found in different accents [20] American, French, Mexican, Indian etc and the use and implementation of the language differs in context. A location-based dictionary can be generated in the future to maintain the diversity in the same language. Hence, a user can natively speak in their accent of choice without actively trying to learn it. A further enhancement of this system can also be implemented in AI-based bots that can respond by interacting natively thereby improving its speech patterns. Collectively this overall enhances the system making it efficient and accurate with more implementations and usage.

## 6. REFERENCES


[1] Oxford Living Dictionaries, How many words are there in the English Language? https://en.oxforddictionaries.com /explore/how-many-words-are-there-in-the-english-language/

[2] Learning Lab 2013, *Academic style*, RMIT, 29 November 2013. http://emedia.rmit.edu.au/learninglab/content/ -academic-style

[3] UniLearning 2000, *Academic writing*, UOW, 29 November 2013, http://unilearning.uow.edu.au/academic/1a.html

[4] UTS, "What is the difference between formal and informal language?" https://www.uts.edu.au/current-students/support/helps/self-help-resources/grammar/formal-and-informal-language

[5] British Council, « Sentence Structure » https://learnenglish.britishcouncil.org/en/english-grammar/sentence-structure

[6] Machine Learning, https://en.wikipedia.org/wiki/Machine_learning

[7] Zipf's Law, https://simple.wikipedia.org/wiki/Zipf%27s_law

[8] Decision Tree, https://en.wikipedia.org/wiki/Decision_tree_learning

[9] Language Model, https://en.wikipedia.org/wiki/Language_model

[10] Karpathy, Andrej. "The Unreasonable Effectiveness of Recurrent Neural Networks".Bengio, Yoshua (2008). "Neural net language models". Scholarpedia. 3. p. 3881.

[11] Mikolov, Tomas; Chen, Kai; Corrado, Greg; Dean, Jeffrey (2013). "Efficient estimation of word representations in vector space". arXiv:1301.3781 Freely accessible.

[12] Mikolov, Tomas; Sutskever, Ilya; Chen, Kai; Corrado, Greg S.; Dean, Jeff (2013). Distributed Representations of Words and Phrases and their Compositionality (PDF). Advances in Neural Information Processing Systems. pp. 3111–3119.

[13] Harris, Derrick (16 August 2013). "We're on the cusp of deep learning for the masses. You can thank Google later". Gigaom.

[14] Weights and Biases, https://www.packtpub.com/mapt/book/big_data_and_business_intelligence/9781788397872/1/ch01lvl1sec12/weights-and-biases

[15] Artificial Neural Networks, https://en.wikipedia.org/wiki/Artificial_neural_network

[16] Learning to predict by the methods of temporal differences. RS Sutton - Machine learning, 1988 – Springer

[17] Recurrent neural network-based language model T Mikolov, M Karafiát, L Burget, J Černocký… - … Annual Conference of …, 2010 - isca-speech.org

[18] Content-based recommendation systems MJ Pazzani, D Billsus - The adaptive web, 2007 – Springer

[19] Deep learning Y LeCun, Y Bengio, G Hinton - nature, 2015 - nature.com


 