



Text Prediction in Notepad Using Machine Learning Techniques

¹Lavanya Mathiyalagi A, ²Chellameena S

¹Assistant professor, Computer Science and Engineering, Einstein College of Engineering, Tirunelveli.

²Student, Computer Science and Engineering, Einstein College of Engineering, Tirunelveli.

ABSTRACT:

Information technology plays a awfully important role in society. People with disabilities are often limited by slow text input speed despite the employment of assistive devices. Word prediction aims at easing the text entry by offering the subsequent word by suggesting the list of most probable words. Word prediction is vital phenomena in typing that benefit users who type using keyboard or other similar devices. However, new applications, like writing short text messages via mobile phones, have recently appeared. An unlimited amount of heterogeneous text prediction methods and techniques are often found in literature. This project aims to make notepad and connected to database inserting data and retrieving data from notepad and also implements speech to text this plays easy to type and plays vital role in computer field.

Index terms: Machine learning, Text Prediction, Speech recognition, JavaFX, LSTM algorithm, MySQL database.

1. INTRODUCTION

Text prediction is one in all the foremost widely used techniques to boost the communication rate in augmentative and alternative communication. Prediction systems are traditionally utilized by people with disabilities (e.g., people with motor and speech impairments). However, new applications, like writing short text messages via mobile phones, have recently appeared. An unlimited amount of heterogeneous text prediction methods and techniques is found in literature. Their heterogeneity makes it difficult to grasp and compare them, so as to pick out the foremost convenient technique for a selected design. Word prediction is an “intelligent” word processing feature that can alleviate writing breakdowns for a range of students simply by reducing the number of keystrokes necessary for typing words.

Word prediction works like, as a student types, the software monitors the input letter-by-letter, and produces a list of words beginning with the letter sequence recorded. Each time a letter is added, the list is updated. When the target word appears in the list, it can be chosen and inserted into the ongoing text with a single keystroke. (Typically word lists are numbered and words can be chosen by typing the corresponding number). The choice of which predictive text system is the best to use involves matching the user’s preferred interface style, the user’s level of learned ability to operate predictive text software, and the user’s efficiency goal. There are various levels of risk in predictive text systems, versus multi-tap systems, because the predicted text that is automatically written that provide the speed and mechanical efficiency benefit, could, if the user is not careful to review, result in transmitting misinformation. Predictive text systems take time to learn to use well, and so generally, a device’s system has user options to set up the choice of multi-tap or of any one of several schools of predictive text methods. Word prediction can help a student during word processing by “predicting” a word the student starts to type and providing spoken feedback. When a student inputs the first letter of a word, the software presents a list of possible words beginning with that letter. As each additional letter is added, the list is refined. When the intended word appears in the list, the student selects it, often by clicking on it or typing its number, which inserts the word into the document. Auditory support is often provided for students who have difficulty reading words presented in the list.

1.1 MACHINE LEARNING

Machine learning (ML) could be a variety of computer science that permits software applications to become more accurate at predicting outcomes without being explicitly programmed to try and do so. Machine learning algorithms use historical data as input to predict new output values. Machine learning could be a field of inquiry dedicated to understanding and building methods that ‘learn’, that is, methods that leverage data to boost performance on some set of tasks. It’s seen as a component of AI. Machine learning algorithms build a model supported sample data, referred to as training data, so as to create predictions or decisions without being explicitly programmed to try to to so. Machine learning algorithms are utilized in a good form of applications, like in medicine, email filtering, speech recognition, and computer vision, where it’s difficult or unfeasible to develop conventional algorithms to perform the needed tasks. A subset of machine learning is closely associated with computational statistics, which focuses on making predictions using computers, but not all machine learning is statistical learning. Some implementations of machine learning use data and neural networks in a very way that mimics the working of a biological brain. In its application across business problems, machine learning is additionally said as predictive analytics. A core objective of a learner is to generalize from its experience. Generalization during this context is that the ability of a learning machine to perform accurately on new, unseen examples/tasks after having experienced a learning data set. The training examples come from some

generally unknown probability distribution (considered representative of the space of occurrences) and also the learner should build a general model about this space that permits it to supply sufficiently accurate predictions in new cases.

1.2 MACHINE LEARNING IN TEXT PREDICTION

Prediction in machine learning refers to the output of an algorithm after it's been trained on a historical dataset and applied to new data when forecasting the likelihood of a specific outcome. Machine learning increases the speed at which data is processed and analyzed. With machine learning, predictive analytics algorithms can train on even larger data sets and perform deeper analysis on multiple variables with minor changes in deployment. Predictive modeling may be a technique that uses mathematical and computational methods to predict a happening or outcome. A mathematical approach uses an equation-based model that describes the phenomenon into consideration. The model is employed to forecast an outcome at some future state or time based upon changes to the model inputs. The model parameters help explain how model inputs influence the end result. The computational predictive modeling approach differs from the mathematical approach because it relies on models that aren't easy to elucidate in equation form and infrequently require simulation techniques to make a prediction.

1.3 MACHINE LEARNING IN SPEECH RECOGNITION

Machine learning consists of supervised and unsupervised learning among which supervised learning is employed for the speech recognition objectives. Supervised learning is that the data processing task of inferring a function from labeled training data. Speech recognition is that the current trend that has gained focus over the decades. Most automation technologies use speech and speech recognition for various perspectives. This paper demonstrates an summary of major technological standpoint and gratitude of the elementary development of speech recognition and provides impression method has been developed in every stage of speech recognition using supervised learning. The project will use DNN to acknowledge speeches using magnitudes with large datasets. Speech Recognition that otherwise referred to as automatic speech recognition (ASR) distinguishes the spoken words, persons and transfers them to a machine-readable configure.

By transferring spoken audio into text, speech recognition skill enhances the users to arrange digital devices. The standard tools like keystrokes, buttons, keyboards etc. In recent years, neural networks of deep learning have taken force within the field of artificial sensory development like machine vision. Within the emulation of this sense, the developments evolve from the segmentation of objects into images and localization of elements, e.g., localization of pedestrians where more recently, these techniques of usage of neural networks are applied to voice recognition. The initial developments that made use of neural networks in speech recognition began to implement differing kinds of basic networks like time-delay neural networks. Because of the low processing capacity that computers had within the 90's, the neural networks didn't have great depth, however, due to the progress in processing speed, it had been possible to begin deploying ever deeper neural networks, making them increasingly efficient in pattern recognition for this reason the interest has arisen within the application of those in tasks of speech recognition, even above other recognition techniques. The most objective of speech recognition is just too capable of understand the spoken information. The related methods failed to execute the expected accuracy, therefore the proposed methodology has been constructed to implement the speech recognition in an automatic manner.

1.4 JAVA FX

JavaFX may be a software platform for creating and delivering desktop applications, likewise as rich Internet applications (RIAs) which will ran into a large style of devices. JavaFX is meant to switch Swing because the standard GUI library for Java SE, but both are going to be included for the foreseeable future. JavaFX has support for desktop computers and web browsers on Microsoft Windows, Linux, and macOS. On desktops, JavaFX supports Windows Vista, Windows 7, Windows 8, Windows 10, macOS and Linux operating systems. Beginning with JavaFX 1.2, Oracle has released beta versions for OpenSolaris. On mobile, JavaFX Mobile 1.x is capable of running on multiple mobile operating systems, including Symbian OS, Windows Mobile, and proprietary real-time operating systems. With the discharge of JDK 11 in 2018, Oracle has made JavaFX a part of the OpenJDK under the OpenJFX project, so as to extend the pace of its development. Oracle support for JavaFX is additionally available, for this long-term version (Java JDK 8), through March 2025. Open-source JavaFXPorts works for IOS (iPhone and iPad) and Android and embedded (Raspberry Pi); and therefore the related commercial software created under the name "Gluon" supports the identical mobile platforms with additional features plus desktop. This enables one ASCII text file base to form applications for the desktop, IOS, and Android.

Features

JavaFX 1.1 was supported the concept of a "common profile" that's intended to span across all devices supported by JavaFX. This approach makes it possible for developers to use a typical programming model while building an application targeted for both desktop and mobile devices and to share much of the code, graphics assets and content between desktop and mobile versions. To handle the requirement for tuning applications on a selected class of devices, the JavaFX 1.1 platform includes APIs that are desktop or mobile-specific. As an example, the JavaFX Desktop profile includes Swing and advanced visual effects. For the tip user, the "Drag-to-Install" feature enables them to pull a JavaFX widget – an application residing during a website – and drop it onto their desktop. The applying won't lose its state or context even after the browser is closed. An application may be re-launched by clicking on a shortcut that gets created automatically on the user's desktop. This behavior is enabled out-of-the-box by the Java applet mechanism since the Java 6u10 update, and is leveraged by JavaFX from the underlying Java layer. Sun touts "Drag-to-Install" as opening from a brand new distribution model and allowing developers to "break aloof from the browser". JavaFX 1.x included a collection of plug-ins for Adobe Photoshop and Illustrator that enable advanced graphics to be integrated directly into JavaFX applications.

The plug-ins generate JavaFX Script code that preserves the layers and structure of the graphics. Developers can then add animation or effects to the static graphics imported. There's also an SVG graphics converter tool (also called Media Factory) that enables for importing graphics and previewing assets after the conversion to JavaFX format. Before version 2.0 of JavaFX, developers used a statically typed, declarative language called JavaFX Script to create JavaFX applications. Because JavaFX Script was compiled to Java byte code, programmers could also use Java code instead. JavaFX applications could run on any desktop that might run Java SE or on any portable that might run Java ME. JavaFX 2.0 and later is implemented as a "native" Java library, and applications using JavaFX are written in "native" Java code. JavaFX Script has been scrapped by Oracle, but development is being continued within the Visage project. JavaFX 2.x doesn't support the Solaris OS or mobile phones; however, Oracle plans to integrate JavaFX to Java SE Embedded 8, and Java FX for ARM processors is in developer preview phase. Sun Microsystems licensed a custom typeface called Amble to be used on JavaFX-powered devices. The font family was designed by mobile computer program design specialists Punch cut and is out there as a part of the JavaFX SDK 1.3 Release. JavaFX could be a tool to style user interfaces for java, it came to exchange Swing during this area. Personally one amongst the most advances that it provides, was changing its syntax to the Java language. This manner it makes learning and using JavaFX very easy for Java developers.

MYSQL DATABASE

MySQL is an open-source electronic database management system (RDBMS). Its name may be a combination of "My", the name of co-founder Michael Widenius's daughter, and "SQL", the abbreviation for Structured command language. A electronic database organizes data into one or more data tables within which data types is also associated with each other; these relations help structure the information. SQL may be a language programmers use to form, modify and extract data from the on-line database, in addition as control user access to the database. Additionally to relational databases and SQL, an RDBMS like MySQL works with an software package to implement a electronic database in a very computer's storage system, manages users, allows for network access and facilitates testing database integrity and creation of backups. MySQL is free and open-source software under the terms of the GNU General Public License, and is additionally available under a spread of proprietary licenses.

1.5 ML MATCHING

Natural Language Processing and Algorithmic Probability are used in Machine Learning. The system reads the entire user input and analyzes it thoroughly. The strength of the match is determined by the user's confidence score configuration. The default matching system is ML, which is turned on by default.

1.6 PROBLEM STATEMENT

Word prediction is vital phenomena in typing that benefit users who type using keyboard or other similar devices. In this project visiting create notepad and connected to database inserting data to that and retrieving data from notepad this plays easy to type and plays vital role in computer field. This is the notepad with special features which enhance the performance of the application. so the feature includes speech to text recognition technology which helps to use the application more efficiently. By using this technology user can really save the time and also the manpower. Speech to text is a speech recognition software that enables the recognition and translation of spoken language into text through computational linguistics

2. SYSTEM STUDY

2.1 EXISTING SYSTEM

Next word prediction is one of NLP fields because the LSTM model to make the prediction with 200 epochs. The result showed that it maintained to get accuracy 75% while the loss was 55%. Based on that result, it could be said the accuracy is good enough. Other researches which used different models. The model could be used to predict the next word by giving the input of the destination. To make a model to predict the next word using LSTM. LSTM which stands for Long Short Term Memory is a variant of the recurrent neural network (RNN) architecture. The reason why the researcher used this model is because it suits this case as it can have a longer memory of what words are important. The aim of creating this model is to predict the next word based on the input which the result should predict correctly. The researcher also would like to know how accurate it's.

2.1.1 Disadvantages of Existing System

This methodology produce 75% accuracy on text prediction. This makes the prediction speed slow. So user can't use the notepad in the easy way. It produce some confusions on the data.

2.2 PROPOSED SYSTEM

The proposed system analyses the next word prediction methodologies, the notepad with special features which enhance the performance of the application. so the feature includes text prediction and speech to text recognition technology which helps to use the application more efficiently. By using this technology user can really save the time and also the manpower. Speech to text is a speech recognition software that enables the recognition and translation of spoken language into text through computational linguistics. In the proposed system we have to examine the LSTM algorithm for text

prediction. For analysis the text prediction algorithms ,the accuracy, Block size, Prediction applications and related features, dictionary structure, prediction method, user interface, etc., are examined.

2.2.1 Advantages of Proposed System

The proposed system provides the advanced notepad which reduce the man work. The notepad contains special features like text prediction, speech recognition which helps easy to type large documents, and also used for take a quick notes. The proposed system try to increase accuracy level and speed using high quality database.

3. RESULTS AND DISCUSSION

This chapter discusses the overview of Text prediction in notepad using machine learning and also speech recognition techniques in machine learning, the input wss given by the user.

3.1 Overview

Predictive text is an input technology used where one key or button represents many letters, like on the numeric keypads of mobile phones and in accessibility technologies. Each key press ends up in a prediction instead of repeatedly sequencing through the identical group of “letters” it represents, within the same, invariable order. Predictive text could leave a whole word to be input by single key press. Predictive text makes efficient use of fewer device keys to input writing into a text message, an e-mail, an address book, a calendar, and also the like. There are many ways to create a tool that predicts text, but all predictive text systems have initial linguistic settings that supply predictions that are re-prioritized to adapt to every user. This learning adapts, by way of the device memory, to a user’s disambiguating feedback that ends up in corrective key presses, like pressing a “next” key to urge to the intention. Most predictive text systems have a user database to facilitate this process. This is often approximately true providing that every one words used are in its database, punctuation is ignored, and no input mistakes are made typing or spelling.

3.2 Notepad with Text prediction

- 1) Creating Notepad
- 2) Connecting Database
- 3) Inserting Data in Database
- 4) Retrieving Data from database

3.3 Notepad with speech recognition

Speech to text could be a speech recognition that allows the popularity and translation of voice communication into text through linguistics. It’s also referred to as speech recognition or computer speech recognition. Specific applications, tools, and devices can transcribe audio streams in real-time to display text and act thereon. speech recognition systems depend upon software and services to function adequately, with the most type being built-in dictation technology. Many devices now have built-in dictation tools, like laptops, smartphones, and tablets like all varieties of technology, speech to text has many benefits that help us improve daily processes. These are a number of the most advantages of using speech to text

- Save time: Automatic speech recognition technology saves time by delivering accurate transcripts in real-time.
- Cost-efficient: Most speech to text software features a subscription fee, and some services are free. However, the value of the subscription is much more cost-efficient than hiring human transcription service is the source of this event. Each of the Result Tokens in the array that the get Best Tokens method produces corresponds to a particular spoken phrase.

4. CONCLUSION AND FUTURE SCOPE

There are multiple automatic text summarizers with great capabilities and giving good results. We have learned all the basics of Extractive and Abstractive Method of automatic text summarization and tried to implement extractive one. We have made a basic automatic text summarizer using nltk library using java and it is working on small documents. We have used extractive approach to do text summarization. We have successfully implemented state-of-the-art model for abstractive sentence summarization to recurrent neural network architecture. The model is a simplified version of the encoder-decoder framework for machine translation.

4.1 FUTURE SCOPE

- The notepad with language translation using text prediction.
- Text prediction and speech recognition with more accuracy level..

References

-
- [1]. HongLuo, "A new automatic prediction system of optical machine learning algorithms for text classification", iee , international conference, 2019.
 - [2]. Wei song, "A Text generation and prediction system:pre-training on new corpora using Bert and gpt2",iee,10th international conference,2020.
 - [3]. V.D.Ambreth Kumar, "Recurrent Neural Network based Text Summarization Techniques by Word Sequence Generation", 6th international conference, 2021.
 - [4]. Au guo, "Automatic Prediction and Insertion of Multiple Emojis in Social Media Text", international conference, 2020.
 - [5]. Varsha verma, "A review on machine learning based student's academic performance prediction systems", international conference,2020.
 - [6]. Ashokshah, "Machine Learning-based Predictive Analytics of Student Academic Performance in STEM Education", iee global engineering conferences,2019.
 - [7]. Ryuchi Nisimura, "Speech-to-text input method for web system using JavaScript", international conference, 2018.