MULTI LAYER NEURAL NETWORK AND SOUNDEX ALGORITHMS FOR TEXT RECOGNITION MODEL

Mandeep Kaur
Department Of Computer Science and Engineering
CT Institute of Engineering & Technology
Jalandhar Punjab India
man06deep@gmail.com

Sheetal Mahajan
Department Of Computer Science and Engineering
CT Institute of Engineering & Technology
Jalandhar Punjab India
Sheetalmahajan57@gmail.com

Abstract— This paper presents a text recognition model based on two algorithms multi-layer neural networks and Soundex algorithms. The text recognition model receives, recognize the text SMS sent by the user. Based on recognition, text recognition model search data from the database and thus provides the appropriate files to the user via Email. To increase the matching rate of the designed text SMS, the two classifications are used which improves the classification accuracy.

Keywords-Text recognition Model, Artificial Neural Network, soundex algorithm

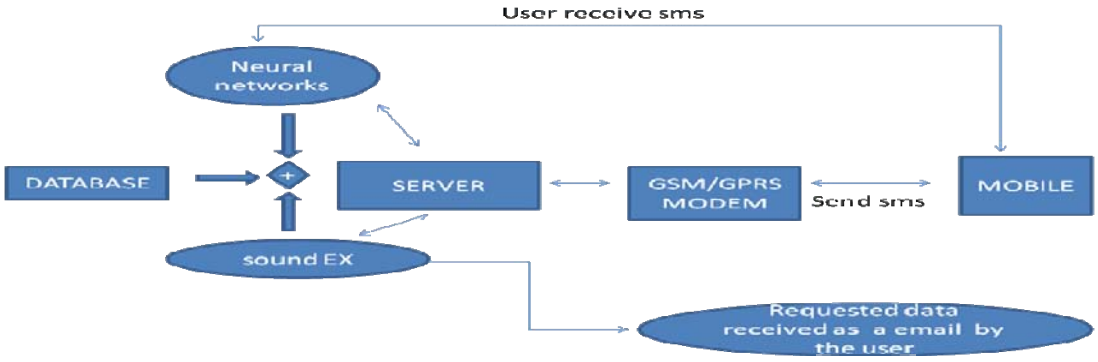
I. INTRODUCTION

The dawn of human civilizations man has been creating things to meet his needs. There is lot of new means of communications which have made our life so simple and easy going. Mobile phones can be seen as the most used present day phenomenon in this fast growing time. Mobile phones are a revolutionary step in the field of telecommunication. In addition to telephony, modern mobile phones also support a wide variety of other services such as SMS, MMS, email, internet access, etc.

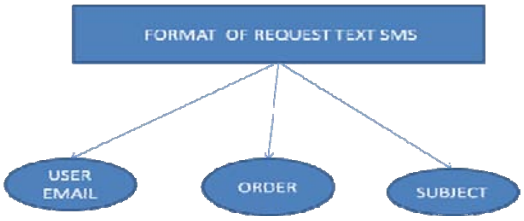
Short message service is a mechanism of delivery of short messages over the mobile networks. SMS was created during the late 1980s to work with digital technology called GSM (Global system for mobile communication) which is the basis for most modern cell phones. SMS is most widely used application in the world with 3.5 billion active users, or 78% of all mobile phones subscribers. The term SMS is used for all types of short text messaging and the user activity itself in the many parts of the world. SMS is also employed in direct marketing known as SMS marketing. The main advantage of SMS is that it is more discreet than a phone conversation, making it the ideal form for communicating when you don't want to be overheard. It is often less-time consuming to send a text message than to make a phone call or send an email. Also, it has become essential for librarian users to access required materials like books, journals, magazines, previous examination papers, novels, etc. A text recognition model is developed to fulfill the user requirements. The user sends an SMS (email-id, order and subject) to a mobile phone at the other end. At the other end, GSM/GPRS modem is installed which receives the SMS and sends data to the computer. Then, multilayer neural networks and soundex algorithms works to recognize the received SMS and thus the required data are sent to the user via Email and SMS.

THE PAPER BASED ON TEXT RECOGNITION MODEL CAN BE ORGANIZED INTO THE FOLLOWING SECTIONS:

1. An overview of text recognition model based on text SMS recognition system:
   The text recognition model works as follows:
   The user sends the request to the GSM/GPRS modem for accessing the data through SMS. The system server receives the text message from GSM modem. The received text message will be processed by a text recognition model to understand the user request. For the recognition process, two algorithms such as artificial neural networks using back propagation algorithm and soundex algorithm are used. If neural networks were unable to recognize the requested data due to any sort of error then the recognition is done by the soundex algorithm. If both algorithms are failed to recognize the request, the message will be sent at the user side of failed operation due to an error message. System after the successful recognition of the text request for the recognized words is to be searched from system for the needed data.



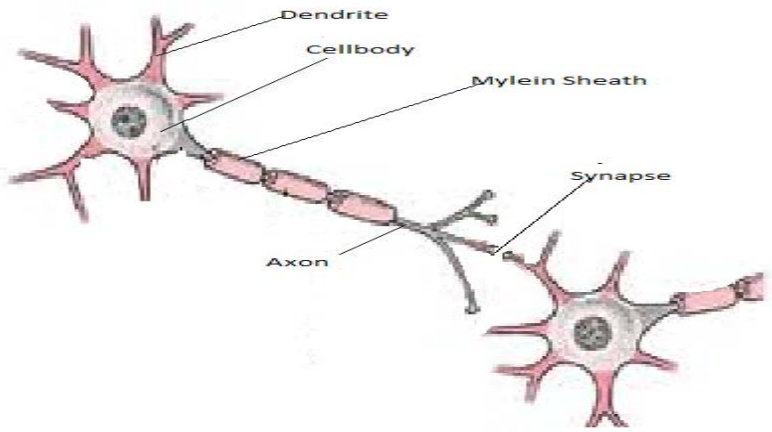
Example of text recognition model:
Suppose the librarian user request for data such as books, journals, magazines, previous examination papers, novels related to various fields such as Network security, Multimedia system, Expert system, etc. for such a request the required request text message format include three things: user email, the order and the subject as shown in figure.



1. The order part of the message include the list of items as required by the librarian such as books, journals, magazines, previous examination papers, novels.
2. The subject part of the message includes the various subjects as related to the requested items.
   Example of request message
   Ritusharma5@gmail.com,journals,computerscience
2. To increase the matching rate of design text SMS system, the combination of multi-layer neural networks and soundex algorithms are used.

2.1 Artificial neural network:

Humans and other animals process information with neural network. Animals are able to react adaptively to changes in their external and internal environments and they use their nervous system to perform these behaviors. An appropriate model of the nervous system should be able to produce similar responses and behaviors in artificial systems.



Artificial neural networks are composed of a set of neurons joined together by synapses. Neuron perform simple task basically yes/no decisions. A synapse is an object which links one neuron connected to its output.

The neurons structure consists of a cell body, many branched dendrites receiving sensory information and a single axon that transfers motor information away from the neuron.

Dendrites are small, branched extensions originating for the cell body which receives signals from other neurons.

Axons, which are longer than the dendrites, originate from the cell body. Each axon is composed of cone-shaped base called axon hillock.

Information is transmitted to the receiving cell at these junctions via. Chemical messenger called neurotransmitters.

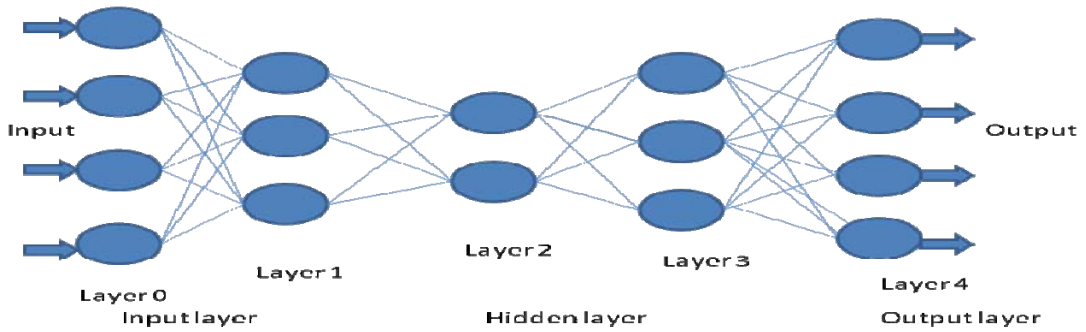
At the end, axon is divided into many branches from the signal passing through the axon is transmitted to the next cell via a junction known as synapse.

Feed forward multi-layer neural network:

A feed forward neural network is an artificial neural network where the connections between units do not form a directed cycle. The feed forward neural network was the first and simplest type of artificial neural network. In this network the information moves in only one direction, forward, from the input nodes, through the hidden nodes and to the output nodes. There are no cycles or loops in the network.

The feed forward neural network has the following characteristics:

1. Perceptrons are arranged in layers, with the first layer taking in inputs and the last layer producing outputs. The middle layer has no connection with the external world and hence is called hidden layers.
2. Each perceptrons in one layer are connected to every perceptron on the next layer and hence information is constantly “fed forward” from one layer to the next, and that’s why these networks are called feed forward networks.
3. There is no connection among perceptrons in the same layer.



Neural networks do recognition of words on the basis of similarities. They are not working correctly where the similarities in the words are not found. Neural networks store pattern information and then superimpose this pattern information on the many synaptic connections between neurons.

2.2 Soundex algorithm:

The term “soundex” algorithm was first developed by Robert C. Russell in 1918. The algorithm converts a surname into a code consisting of the first letter of the surname, followed by three digits. The digits are assigned according to a predetermined grouping of consonants, where the consonants groups share phonetic features. The key concept behind soundex is that the relationship between letters and sound should assure by similar sounding names are assigned the same code.

The soundex algorithm works as follows:

1. Capitalize all letters in the word and drop all punctuation marks. Pad the word with right most blanks as needed during each procedure step.
2. Retain the first letter of the word.
3. Change all occurrences of the following letters to ‘0’ (Zero). ‘A’,’E’,’I’,’O’,’U’,’H’,’W’,’Y’.
4. Change letters from the following sets into the digit given.

Letter	Code
B,F,P,V	1
C,G,J,K,Q,S,X,Z	2
D,T	3
L	4
M,N	5
R	6
H,Y,W	(OMITTED)
A,E,I,O,U	(OMITTED)

5. Remove all pairs of digits which occur beside each other from the string that resulted after step 4.
6. Remove all zeros from the string that results from step 5.
7. Pad the string that resulted from step 6 with zeros and return only the first 4 positions, which will be of the form <uppercase letter> <digit> <digit> <digit>.

Example: Suppose the name “PATER” is sent by the user as “PATRN”. According to the soundex algorithm, the word “PATRN” will have the phonetic value equals to “P365”. This word does not exist in the soundex database. Thus, the algorithm will compare it with the similar words included in the database. Similar words that start with letter P are listed below:

WORDS	PHONETIC VALUE
PATER	P360
PEDDLE	P334
PLANT	P453
PALLAVI	P441

The minimum difference between P365 and the above listed words is associated with word “PATER”. Thus the soundex algorithm will recognize the word “PATRN” as “PATER”.

Conclusion

It is concluded that from the above proposed text recognition model, the neural network was able to recognize the text message correctly when the words have 40% and fewer errors. If the number of errors increases by 40% then soundex algorithm will provide better results for the recognition of the words. Thus, the combination of both the algorithms plays a very good role in enhancing the recognition rate of the received text message.

References

[1] D. Suozil et al. / Chemo metrics and Intelligent Laboratory Systems 39 (1997) 43-62
[2] Artificial neural network by Anil k Jain Michigan state university and Jianchang Mao.
[3] Introduction to artificial neural networks-eduGl.net.
[4] Improving precision and recall for soundex retrieval by David Holmes and Catherine
[5] Mccabe.