

Figure 3: Neural network processing ^[10]

In the proposed system the inputs are the errors and weights are the suitable solution needed to be implemented on the sentences to make it error free. The processing element is the NLP that processes the error and changes the data throughout the chunk of data. The Neurons undergo simulation of various aspects of error in the data and learns the situation that need to be applied on the errors. The processing is done only by NLP with the various condition and the actual output is the correct statement with maximum correction. The artificial neural network process the errors efficiently and rectifies the raw text to valuable information.

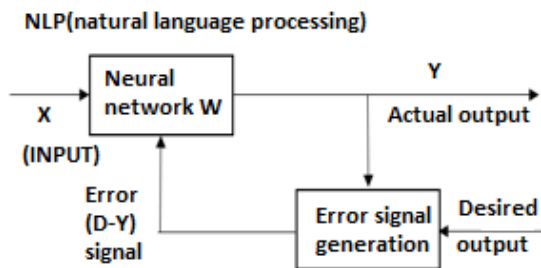


Figure 4: Supervised learning ^[10]

A multivalued solution for a problem can also be considered where the similar errors are taken and the suggested solution is set as the bias which are applied on the errors to get the suitable statement without changing the meaning and the weights. For this NLPbased suitable solution is used in rectifying errors. The outer layer of neurons process the node with each bias set and the NLP takes the decision adapting to the suitable solution for each error and corrects it. This method of correcting is carried out under the process of reinforcement learning using unsupervised learning.

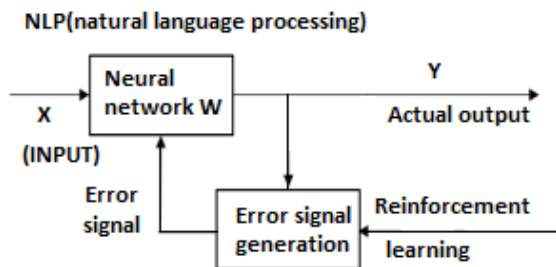


Figure 5: Reinforcement learning (unsupervised learning) ^[10]

Reinforcement learning is similar to the supervised learning. In supervised learning, the correct target output values are known for each input pattern. In some cases, less information might be available, or the NLP can frame the statement exactly but the solutions in the learning data may be 50% correct to get the actual output or so thus the critic

information is available not the exact information. The Neural network process the error input with various learning data and try to get the correct text which is the actual output. The learning based on the critic information is called reinforcement learning and the feedback is called reinforcement signal.

The reinforcement learning sub-discipline of unsupervised learning, is used mainly when the solution is not accurate or the user lets the system to take correct solution from the training data and produce quality information. Most often the error are identified by the NLP and suggested with better solution, but if the solution is not appropriate then the active user can input the actual solution. The user can also provide multivalued solution to the system. The system initially works with lot of errors, the neural network gets trained with such errors and develop the training data. As the training data increases the network uses the training data usually on similar problems and further applies reinforcement learning to produce the quality in text.

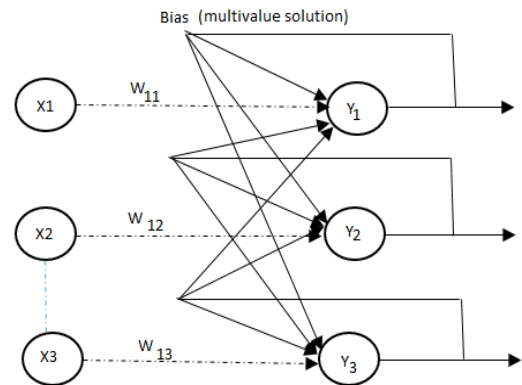


Figure 6: Single layer recurrent network ^[10]

The similar errors are taken as input to the network ($X_1, X_2 \dots X_i$) where the weights ($W_{11}, W_{12} \dots W_{ij}$) are the suggested value by the NLP and the bias is the various multivalued solution to be applied on the errors of similar type. The neurons ($Y_1, Y_2 \dots Y_j$) are the network nodes that work as NLP provides output if the text is not filtered or corrected then error is generated and the statement is again put forth as errors; completely with description to active user to take decision by considering a detailed look through and provide a better solution. The error are feed as input to the input layer, the errors that are similar are being processed if the error are just single in count a single neuron itself can do the processing and generate the expected text, as the count increases layer of input to the network gets complex. The network is single layer recurrent network where the errors are processed of similar type the NLP suggestion is used as the initially to generate the text, if the user have suggested a single value then it takes the user's solution and gets the desired text. If there are multivalued then each value is used in recurrent network and the accuracy of the statement is evaluated and the best value is used to filter the text in the complete chunk. A similar process is carried out for the errors encountered and accepted as multivalued input; by setting each as bias and further evaluated to get the appropriate statement. The NLP plays an important role in evaluating the value to statement and getting the rich text. Even after completion of learning data, if the reinforcement

learning is not successful then the system takes a suitable input for the individual error and corrects as per user's suggestion.

This process of initial phase provides a rich text that is error free and all the human errors are removed to get the maximum accurate information for applying statistical data algorithm. The process of filtering the datasets accurate and perfect as the training data increases, thus the neural network behaves like a human brain and takes the decision as a human brain is capable of taking. The major issues of social networking and E-commerce websites data is eliminated.

Application of this step is, the text data can be classified such as weather data, product review data etc. The NLP while processing the data chunks it can go through the various conditions and process the text to get the data based on various domain. The data can further undergo the common pattern specification or rules based information extraction to extract the information needed for analytics. Association among the entity can be one of the relation to extract the relation data such as any user name, address, state, pin code, contact numbers etc.

3.2 Layout of Proposed System

The working of the system is functional and crucial system is the network. The flow of our system is as follows:

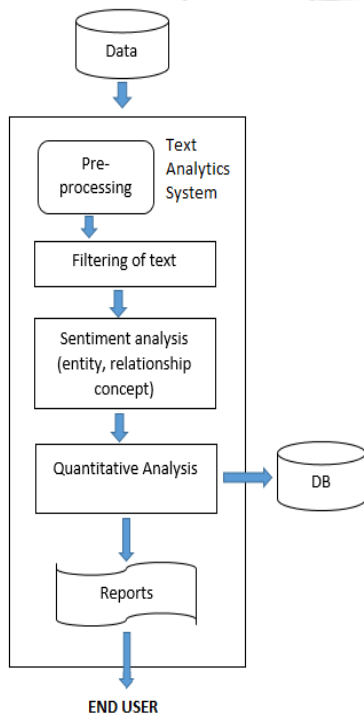


Figure 7: Work flow of the system

The next step in the proposed system is the main mining of the exact data where the sentiment analysis takes place. The sentiment analysis uses the statistical data algorithm i.e. Rule based information retrieval and common pattern specification language where set of rules are being written and the system search that pattern to get the data. The sentiment analysis deals with the entity relationship concept where the user details or the relation information of the particular entity is searched. The pattern are formed based

on the language grammar base i.e. if a word is to be searched from the processed text the various form of it is searched such as its noun, adverb, verb, adjective. The word tends to be similar synonym but the way of using in the sentences differs so many patterns and rules based syntax are written in order to process the text.

The processing of the text by a local system is difficult so a big data tool such as HADOOP, can be used to process the text and get the results faster. The rule based syntax can be written in the mapper and the mapper does the work of searching the text. As the mapper completes its work; the reducer performs the main task of Quantitative analysis. The Reducer combines the count of the words that was given as input to be searched. The system results with a detailed layout of the words in various form and their count.

3.3 System Forms

1	ERRORS	Count	Solution
2	Excellent in functional	80	Error: The sentence formation is wrong
3	its too good	105	"the functionality are grt".
4	Don't purchahse	3	Suggestion: correct stmt can be "The
5	the product is awsm	60	functionality is great", "functionality"
6	the functiionality are grt	12	=> 'Functionality', "grt" => does it mean
7	purchase asap	30	'great'.
8	limited period offer discount %	15	User Suggestion to problem:
9	better purchase some other	5	

Error: Statement error on line : 256 column:75 row: 15.
 "The product functiionality are grt, better than other products in the market".
 Error no : 247 spelling error by the system. kindly suggest the correct meaning or leave on system to resolve [Click here](#)

Figure 8: Error report

The system provides a way to insert the raw data. The data can be any unstructured text form. The data is divided into clusters and each data cluster undergo the operation. The filtering of the text takes place where the human error is shown and the NLP suggest the solution in the form of row/tab. The user can go with NLP suggestion by clicking the self-learning or click here option. The errors undergone neural network operations or the active user can provide a solution for it. As training data increases the system performs the reinforcement learning and the filtration of text; else the errors are shown in the window fig (8).

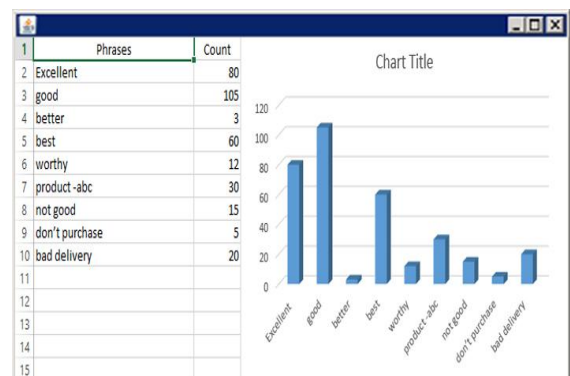


Figure 9: Detailed report

After the completion of the error evaluation the algorithm performs the analysis of the entity using the HADOOP environment. The Quantitative analysis is carried out and the count of the words are shown in form of report. The reports could be used by the organization in decision making. The text analytics of the system is very accurate; the system at the initial stage removes the error with some human intervention and later the neural network behaves like a self-learning system; as most of the errors are evaluated by system itself without waiting for the user to give suggestion as the system itself has the scenario of the previous learning.

4. Application

The proposed system concept can be used in various application. The learning concept in our system can be added to search engines, search engine optimization, social networking applications, E-commerce brand/product reviews and many more. In search engines the user queries the particular text, to get the related information, the system can be implemented similarly where the data is being filtered from the various pages and related links can be shown as well the network maintains its learning data for similar queries. The search engine optimization can also be done as the text that is being filtered or searched more often the URL of the pages can be kept in the memory and when new user searches it once again the user can get the best content related to the queries. The system well suitable for social networking trends and the various customer review about the upcoming event can be known easily by just using the social network data rest all the reports can be modified to some extent as per the requirements.

5. Conclusion

The technology is getting faster with new trends, the amount of data is increasing by leaps and bound from mobiles as compared to the personal computers. The social media, e-commerce is a stream of getting the huge amount of data and filtering the data for complete use is challenging, the proposed system get an accurate rich text for analytics. The other unstructured data such as image, videos, sound etc. filtering and performing analytics is more challenging as image processing and various pattern matching techniques would help to some extent. The various pattern matching techniques used in text mining are efficient in getting the text data and relation among them. The hidden pattern of the text information can be used by the organization in taking a firm decision for the growth and achievement of the goal as well as to know the end user want from the organization.

References

- [1] Dr.Goutam Chakraborty "Analysis of Unstructured Data: Applications of Text Analytics and Sentiment Mining", Paper 1288-2014
- [2] "Integrated Visual Analytics Tool for Heterogeneous Text Data" IEEE Symposium on Visual Analytics Science and Technology October 24 - 29, Salt Lake City, Utah, USA 978-1-4244-9487-3/10/\$26.00 ©2010 IEEE
- [3] Alex Endert, Patrick Fiaux, Chris North" Semantic Interaction for Visual Text Analytics"
- [4] Vishal Gupta, Gurpreet S. Lehal "A Survey of Text Mining techniques and applications, JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE, VOL. 1, NO. 1, AUGUST 2009
- [5] Text Mining Techniques- A Survey "International Journal of Advanced Research in Computer Science and Software Engineering "Volume 2, Issue 4, April 2012 ISSN: 2277 128X
- [6] Big Data and Text Analytics: Solving the Missing "Big Language" Link, Forrester Research, Inc., 2013 Customer Experience Predictions, January 2013
- [7] "Visual Analytics of Text Streams Through Multiple Dynamic Frequency Matrices", IEEE Symposium on Visual Analytics Science and Technology 2014, November 9-14, Paris, France.
- [8] "Integrated Visual Analytics Tool for Heterogeneous Text Data", IEEE Symposium on Visual Analytics Science and Technology 2014, November 9-14, Paris, France.
- [9] "Word Cloud Explorer: Text Analytics based on Word Clouds", 2014 47th Hawaii International Conference on System Science. 978-1-4799-2504-9/14 \$31.00 © 2014 IEEE DOI 10.1109/HICSS.2014.231.
- [10] Book - Principles of Soft computing , Author - Sivanandam and Deepa
- [11] <http://www.ibm.com/software/data/bigdata/>
- [12] <http://www.simafore.com/blog/bid/116340/Keyword-clustering-using-web-mining-and-text-mining-with-RapidMiner> -posted by Bala Deshpande.

Author Profile

Anil Kumar Bhandi: Student of MCA from Vivekanand Education Society's Institute of Technology, Chembur (University of Mumbai). Bachelor's Degree in BScIT (Information Technology) from S.I.W.S College, Wadala (University of Mumbai).

Prof. Shivkumar Goel: Deputy Head of Department MCA in Vivekanand Education Society's Institute of Technology, Chembur (University of Mumbai).