International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 08 | August -2020 ISSN: 2582-3930

Review on Natural Language Processing and its Applications

Rashmi Katratwar
Priyadarshini Bhagwati College of Engg., Nagpur

Prof. Archana A. Nikose
Priyadarshini Bhagwati College of Engg., Nagpur

ABSTRACT: Rapid development in the field of ML, AI or NLP is taking the system at next level in the terms of intelligence and the decision making. On other hand education sector are growing by making different experiments in teaching methodologies but still systems are lacking with teaching to naïve users or teenagers. Searching the image on the basis of its name is easier task rather searching it on the basis of the content was always big challenge. We have different dictionaries where individual words are represented using the statements and not the graphical presentation. Displaying the words through graphics has its own limitation it requires a natural language processing technique to understanding the user input which may be at different aspects and it becomes more difficult if it's in regional languages. This papers mainly focused on application and implementation methodologies of NLP.

Keywords, NLP, segmentation, token, parsing

[1] INTRODUCTION

Natural language processing (NLP) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages and, in particular, concerned with programming computers to fruitfully process large natural language corpora. Challenges in natural language processing frequently involve natural language understanding, natural language generation (frequently from formal, machine-readable logical forms), connecting language and machine perception, managing human-computer dialog systems, or some combination thereof.

Statistical Natural Language Processing

Since the so-called "statistical revolution" in the late 1980s and mid-1990s, much Natural Language Processing research has relied heavily on machine learning. Formerly, many language-processing tasks typically involved the direct hand coding of rules, which is not in general robust to natural language variation. The machine-learning paradigm calls instead for using statistical inference to automatically learn such rules through the analysis of large *corpora* of typical real-world examples (a *corpus* (plural, "corpora") is a set of documents, possibly with human or computer annotations).

Many different classes of machine learning algorithms have been applied to NLP tasks. These algorithms take as input a large set of "features" that are generated from the input data. Some of the earliest-used algorithms, such as decision trees, produced systems of hard if-then rules similar to the systems of hand-written rules that were then common. Increasingly, however, research has focused on statistical models, which make soft, probabilistic decisions based on attaching real-valued weights to each input feature. Such models have the advantage that they can express the relative certainty of many different possible answers rather than only one, producing more reliable results when such a model is included as a component of a larger system. Systems based on machine-learning and NLP algorithms have many advantages over hand-produced rules:

- The learning procedures used during machine learning automatically focus on the most common cases, whereas when writing rules by hand it is often not at all obvious where the effort should be directed.
- Automatic learning procedures can make use of statistical inference algorithms to produce models that are robust to unfamiliar input (e.g. containing words or structures that have not been seen before) and to erroneous input (e.g. with misspelled words or words accidentally omitted). Generally, handling such input gracefully with

hand-written rules—or more generally, creating systems of hand-written rules that make soft decisions—is extremely difficult, error-prone and time-consuming.

- Systems based on automatically learning the rules can be made more accurate simply by supplying more input data. However, systems based on hand-written rules can only be made more accurate by increasing the complexity of the rules, which is a much more difficult task. In particular, there is a limit to the complexity of systems based on hand-crafted rules, beyond which the systems become more and more unmanageable. However, creating more data to input to machine-learning systems simply requires a corresponding increase in the number of man-hours worked, generally without significant increases in the complexity of the annotation process.

[2] RELATED WORK

The following is a list of some of the most commonly researched tasks in NLP. Note that some of these tasks have direct real-world applications, while others more commonly serve as subtasks that are used to aid in solving larger tasks. Though NLP tasks are obviously very closely intertwined, they are frequently, for convenience, subdivided into categories; a coarse division is given below.

NLP has different implementation while some of them are listed below,

1. **Sentiment Analysis:** Mostly used on the web & social media monitoring, Natural Language Processing is a great tool to comprehend and analyse the responses to the business messages published on social media platforms. It helps to analyse the attitude and emotional state of the writer (person commenting/engaging with posts). This application is also known as opinion mining. It is implemented through a combination of Natural Language Processing and statistics by assigning values to the text (positive, negative or neutral) and in turn making efforts to identify the underlying mood of the context (happy, sad, angry, annoyed, etc.)

2. Chatbots: We hear a lot about Chatbots these days, Chatbots are the solution for consumer frustration regarding customer care call assistance. They provide modern-day virtual assistance for simple problems of the customer and offload low-priority, high turnover tasks which require no skill. Intelligent Chatbots are going to offer personalized assistance to the customer in the near future**.**

3. Customer Service: Ensuring customer loyalty by keeping them content and happy is the supreme challenge and responsibility of every business organization. NLP has aided in multiple functions of customer service and served as an excellent tool to gain insight into audience tastes, preferences and perceptions. Speech separation where the AI will identify each voice to the corresponding speaker and answer each of the callers separately. An excellent text to speech systems could even aid the blind. For example, a call recording of the customer can give insight into whether the customer is happy or sad, what are their needs and future requirements.

4. Managing the Advertisement Funnel: What does your consumer need? Where is your consumer looking for his or her needs? Natural Language Processing is a great source for intelligent targeting and placement of advertisements in the right place at the right time and for the right audience. Reaching out to the right patron of your product is the ultimate goal for any business. NLP matches the right keywords in the text and helps to hit the right customers. Keyword matching is the simple task of NLP yet highly remunerative for businesses.

5. Market Intelligence: Business markets are influenced and impacted by market knowledge and information exchange between various organizations, stakeholders, governments and regulatory bodies. It is vital to stay up to date with industry trends and changing standards. NLP is a useful technology to track and monitor the market intelligence reports for and extract the necessary information for businesses to build new strategies. Widely used in financial marketing, NLP gives exhaustive insights into employment changes and status of the market, tender delays, and closings, or extracting information from large repositories.

[3] LITERATURE SURVEY

In [1], In this paper, author propose a device which can understand the documents having

comprehensions, keywords, news, articles, images, handwritten texts and summarize the text
in a few hundred words given 2-3 pages of document. Each of the letters and words are analyzed with Image processing techniques and features such as text/ images are extracted. Fig. 1 shows our proposed device having similar design. Proposed device functions in three different stages.

In [2] proposed method of Content Based Video Retrieval (CBVR) the Ordered-Dither Block Truncation Coding (ODBTC) technique is employed which generates appropriate image contents. Combinations of Void-and-cluster half-toning and Block Truncation Coding (BTC) offers low complexity in algorithm and provides better video image quality. Dither array Look-Up-Table (LUT) is a distinctive feature of ODBTC which reduces the difficulties by providing look up values of segmented blocks. ODBTC encoded streams are used for generation of two distinct features including of color features namely Color Co-occurrence
Feature (CCF) and Bit Pattern Features (BPF). After quantizing and bit-mapping from ODBTC encoder, BPF is obtained by LUT. In the presented system, CBVR is achieved by Block Truncation (BT) of expected video information to be retrieved. Proposed system provides good remedy for CBVR for large digital video-data processing in the fields of Image and Video Processing.

In [3] paper discusses a tool for video structure analysis, feature extraction, classification and semantic querying suitable for an extremely broad scale of video data set. The tool analyses the video structure to detect shot boundaries where shots in each video are identified using image duplication techniques. A single frame from each shot is passed to a deep learning model implemented using TensorFlow, that is trained for feature extraction and classification of objects in each frame. Subsequently, an automatic textual annotation is generated for each video and finally with the aid of ontology, semantic searching is done using NLP, which allows receiving an efficient result other than manual video annotation of a large scale dataset. While maintaining accurate querying with automatic video content analysis and annotation with semantic searching with around seventy-four percent accuracy rate, this becomes a useful tool in video tagging and annotation.

In [4] In this study, retrieval techniques based on features like HSV, Color Moment, HSV and Color Moment, Gabor Wavelet and Wavelet Transform, Edge Gradient are studied and implemented. An approach is proposed for retrieval based on combination of color, texture and edge features of image. Performance evaluation of studied image retrieval techniques and proposed technique is done using parameters like Sensitivity, Specificity, Retrieval score, Error rate and Accuracy. Experimental results of performance evaluation demonstrate that proposed technique outperforms other techniques.

In [5] In this paper, we describe a scalable and economical architecture for performing container based parallelization to obtain the best possible quantized image using different quantization techniques on the cloud. This approach using containers can be scaled to be used with huge datasets. The quantization techniques used in this paper are fuzzy entropy and genetic algorithm based techniques. Different types of membership functions are used in each technique to calculate the fuzzy entropy. The best possible quantized image is determined using the Structural Similarity Index (SSIM). This is a futuristic approach for solving lengthy repetitive serial problems in a parallel and economical way. As expected the results significantly better than the serial approach.

[4] METHODOLOGY

In order to have brief idea about NLP and its application we have taken a visualizer application as a example where the system will be trained and programmed to show images on the basis of given textual information. work is mainly divided in to multiple sections or the module which can be described through figure 1.0 which shows the system process flow.

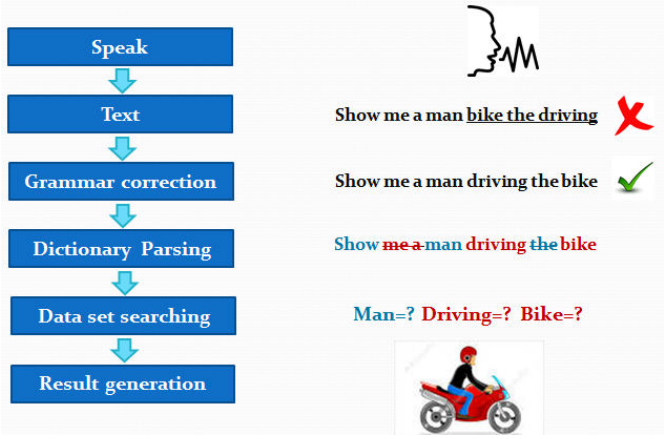

Figure 1.0 System process flow

Main proposed work of the system is,

- Preparing the knowledge dataset and training the system where before system reach at its goal system need to prepare with it's knowledge database. In this step system will be having

option to add new images and defined it's descriptions like header data, searching tags, object available in image, colour shade of the image, etc. this way system will having the information to search.

- Get the user input in the form of speech and convert it in to text possibly in regional language. As the main user base of the system is planned to be a kids or teenagers we are trying the give system speech recognition capability where system will convert speech to text with the help of available speech API or SDK. If available we may use regional language for input.
- Try to find out the grammatical and syntactical errors in the user input and resolve it with the user intervention. It may possible that user input having some syntactical or grammatical mistakes in the statements so first system needs to pre-process the data and make it correct before passing to next step.
- Parsing of the query in order to get the co-related words and removal of unwanted words. Every word in statement having some level of importance so with the help of parsing of input string system will first tokenize the whole input in to multiple tokens and then finds out the most feasible and required words.
- Executing the search query after identifying the activities, object and its properties from user input and final generate the result

CONCLUSION

Natural language processing (NLP) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages. NLP is related to the area of human–computer interaction. Challenges in NLP: natural language understanding, enabling computers to derive meaning from human or natural language input; and others involve natural language generation. Modern NLP algorithms are based on machine learning, especially statistical machine learning. Prior implementations of language-processing tasks typically involved the direct hand coding of large sets of rules. With the help of better use of NLP it has been seen that variety of applications can be implemented with right approach and the techniques. By implementing the proposed system we can take the further steps to make the machine learning and the knowledge set generation graphical at next level.

REFERENCES

1. Soliha Rahman, Kanza Naeem, Dileep Raju Seera, *"Cloud Based Information Retrieval Device for Augmentation of Real Time Images"*, 978-1-5386-1112-8/18/ ©2018 IEEE
2. Savita V. Borkar, Shraddha S. Katariya, *"Content Based Video Retrieval Scheme Using*
3. *Halftone Block Truncation Coding",* 2017 2nd IEEE International Conference On Recent Trends in Electronics Information & Communication Technology (RTEICT), May 19-20, 2017, India
4. Kithmi Ashangani , Wickramasinghe K.U, De Silva D.W.N, Gamwara V.M, Anupiya Nugaliyadde, Yashas Mallawarachchi, *"Semantic Video Search by Automatic Video Annotation using Tensorflow",*Proceedings of the 1st Manufacturing & Industrial Engineering Symposium22 October 2016, Colombo, Sri Lanka
5. Manpreet Kaur, Neelofar Sohi, "*A Novel Technique For Content Based Image*
6. *Retrieval Using Color, Texture And Edge Features"*
7. Mohan Muppidi, Paul Rad, Sos S. Agaian, and Mo Jamshidi, *"Container Based Parallelization for Faster and Reliable Image Segmentation",* 978-1-4799-8633-0/15/ ©2015 IEEE