**LITERATURE REVIEW:**

1. **MURAX: A Robust Linguistic Approach For**

**Question Answering Using An On-Line Encyclopedia**

Author/s: Julian Kupiec

Published Date: 01 July 1993 Source: IEEE

Methodology: Robust linguistic methods are applied to the task of answering closed-class questions using a corpus of natural language. The methods are illustrated in a broad domain:

answering general-knowledge questions using an on-line encyclopedia. A closed-class question is a question stated in natural language, which assumes some definite answer typified by a noun phrase rather than a procedural answer. The methods hypothesize noun phrases that are likely to be the answer, and present the user with relevant text in which they are marked, focussing the user’s attention appropriately. Furthermore, the sentences of matching text that are shown to the user are selected to confirm phrase relations implied by the question, rather than being selected solely on the basis of word frequency. The corpus is accessed via an information retrieval (IR) system that supports boolean search with proximity constraints. Queries are automatically constructed from the phrasal content of the question, and passed to the IR system to find relevant text. Then the relevant text is itself analyzed; noun phrase hypotheses are extracted and new queries are independently made to confirm phrase relations for the various hypotheses.

Pros and Cons: Task evaluation indicates where further effort might be most productive and also indicates how new components contribute to overall performance. The linguistic analysis is based on an underlying regular grammar formalism, both in the HMM tagger and the phrase recognizes. There may be benefits from the use of stochastic context-free grammars, which can also be trained from unlabelled text and enable ambiguity to be quantified in probabilistic terms.

1. **Research on the Hand Gesture Recognition Based**

**on Deep Learning**

Author/s: Jing-Hao Sun, Ting-Ting Ji, Shu-Bin Zhang, Jia-Kui Yang, Guang-Rong Ji

Published Date: 6 Dec. 2018 Source: IEEE

Methodology: With the rapid development of computer vision, the demand for interaction between human and machine is becoming more and more extensive. Since hand gestures are able to express enriched information, the hand gesture recognition is widely used in robot control, intelligent furniture and other aspects. The paper realizes the segmentation of hand gestures by establishing the skin colour model and AdaBoost classifier based on haar according to the particularity of skin colour for hand gestures, as well as the denaturation of hand gestures with one frame of video being cut for analysis. In this regard, the human hand is segmented from the complicated background, the real-time hand gesture tracking is also realized by Cam Shift algorithm. Then, the area of hand gestures which has been detected in real time is recognized by convolutional neural network so as to realize the recognition of 10 common digits. there are many ways to realize hand gesture

segmentation. Based on the segmentation of skin color model, the skin color model is established to realize the hand gesture segmentation according to the difference between skin color of hand gestures and external environment and the model is not affected by the hand postures, but it is not able to exclude the objects which are similar to the skin color, such as human face and so on; the hand gesture segmentation based on edge detection can segment the hand gestures according to the discontinuity of Gray value in the margin area of image region, but it is easy to be interrupted by the noise and it has strict requirements for the background; the hand gesture segmentation based on

movement information, including frame difference method and background difference method and so on adopts the information of movement of hand gestures to segment hand gestures on the premise of static of background. The effect is good in static environment while not well-performed in

dynamic background; the segmentation method of hand gesture based on statistical template matching is able to rapidly identify the hand area and non-hand area by using training classifier of gesture template feature, but it can only recognize one or more hand gestures, it cannot satisfy our

demands. The hand gesture segmentation in the paper pre-processes the images and establishes Gaussian mixture model according to the skin colors, moreover, it also segments hand gestures by combining with AdaBoost classifier based on Haar features.

Pros and Cons: Hand gesture segmentation by using AdaBoost classifier based on Haar feature realizes the acquisition of hand gesture area in complicated environment. Using CamShift algorithm for hand gesture tracking according to the movement of hand gestures and features of deformation ensures to acquire the hand gesture area in real time, finally, the hand gesture area is classified by convolution neural network.

1. **END-TO-END ATTENTION-BASED LARGE VOCABULARY SPEECH RECOGNITION**

Author/s: Dzmitry Bahdanau, Jan Chorowski, Dmitriy Serdyuk, Phil´emon Brakel, Yoshua Bengio

Published Date: 19 May 2016 Source: IEEE

Methodology: The system proposes here is a neural network that can map sequences of speech frames to sequences of characters. While the whole system is differentiable and can be trained directly to perform the task at hand, it can still be divided into different functional parts that work together to learn how to encode the speech signal into a suitable feature representation and to decode this representation into a sequence of characters.

Pros and Cons: The resulting approach is significantly simpler than the dominating HMM-DNN one, with fewer training stages, much fewer auxiliary data and less domain expertise involved. Combined with a trigram language model our system shows decent, although not yet state-of-the-art performance.

1. **SQuAD: 100,000+ Questions for Machine Comprehension of Text**

Author/s: Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang

Published Date: 16 Jun 2016 Source: IEEE

Methodology: Candidate answer generation. For all four methods, rather than considering all O(L2) spans as candidate answers, where L is the number of words in the sentence, we only use spans which are constituents in the constituency parse generated by Stanford CoreNLP. Ignoring punctuation and articles, we find that 77.3% of the correct answers in the development set are constituents. This places an effective ceiling on the accuracy of our methods. During training, when the correct answer of an example is not a constituent, we use the shortest constituent containing the correct answer as the target. SlidingWindow Baseline For each candidate answer, we compute the unigram bigram overlap between the sentence containing it (excluding the candidate itself) and the question. We keep all the candidates that have the maximal overlap. Among these, we select the best

one using the sliding-window approach. Logistic Regression In our logistic regression model, we extract several types of features for each candidate answer. We discretize each continuous feature into 10 equally sized buckets, building a total of 180 million features, most of which are lexicalized features or dependency tree path features.

Pros and Cons: Goal of natural language understanding, we introduce the Stanford Question Answering Dataset, a large reading comprehension dataset on Wikipedia articles with crowdsourced

question-answer pairs. Performance stratified by syntactic divergence. As discussed in Section 4, another challenging aspect of the dataset is the syntactic divergence between the question and answer sentence. shows that the more divergence there is, the lower the performance of the logistic regression model. Interestingly, humans do not seem to be sensitive to syntactic divergence, suggesting that deep understanding is not distracted by superficial differences.

1. **ReadingWikipedia to Answer Open-Domain Questions**

Author/s: Danqi Chen, Adam Fisch, Jason Weston & Antoine Bordes

Published Date: 31 Mar 2017 Source: IEEE

Methodology: Paragraph encoding. Dynamic Coattention Networks. Multi-Perspective Matching.

Pros and Cons: Machine comprehension systems alone cannot solve the overall task. This method

integrates search, distant supervision, and multitask learning to provide an effective complete

system. Evaluating the individual components as well as the full system across multiple benchmarks

showed the efficacy of the approach.

1. **Attention Is All You Need**

Author/s: Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin.

Published Date: 12 Jun 2017 Source: IEEE

Methodology: Most competitive neural sequence transduction models have an encoder-decoder structure.

Here, the encoder maps an input sequence of symbol representations to a sequence

of continuous representations z = (z1; :::; zn). Given z, the decoder then generates an output

sequence (y1; :::; ym) of symbols one element at a time. At each step the model is auto-regressive

, consuming the previously generated symbols as additional input when generating the next.

The Transformer follows this overall architecture using stacked self-attention and point-wise, fully

connected layers for both the encoder and decoder.

Pros and Cons: To evaluate if the Transformer can generalize to other tasks we performed experiments on English constituency parsing. This task presents specific challenges: the output is subject to strong structural constraints and is significantly longer than the input. Furthermore, RNN sequence-to-sequence models have not been able to attain state-of-the-art results in small-data regimes. the first sequence transduction model based entirely on attention, replacing the recurrent layers most commonly used in encoder-decoder architectures with multi-headed self-attention.

1. **Adaptive Document Retrieval for Deep Question Answering**

Author/s: Bernhard Kratzwald, Stefan Feuerriegel

Published Date: 20 Aug 2018 Source: IEEE

Methodology: Threshold-Based Retrieval As a naive baseline, we propose a simple threshold-based heuristic. That is, ni is determined such that the cumulative confidence score reaches a fixed threshold . Ordinal Regression- further implement a trainable classifier in the form of an ordinal ridge regression which is tailored to ranking tasks. It is further expected that cumulative confidence likely to be linear. The classifier then approximates ni with a prediction yi that denotes the position of the first relevant document containing the desired answer.

Pros and cons: establish that deep question answering is subject to a noise information trade-off. As a consequence, the number of selected documents in deep QA should not

be treated as fixed, rather it must be carefully tailored to the QA task. Second, we propose adaptive

schemes that determine the optimal document count. This can considerably bolster the performance of deep QA systems across multiple benchmarks. Third, we further demonstrate how crucial an adaptive document retrieval is in the context of different corpus sizes. Here our adaptive

strategy presents a flexible strategy that can successfully adapt to it and, compared to a fixed document count.

1. **Ranking Paragraphs for Improving Answer Recall in**

**Open-Domain Question Answering**

Author/s: Jinhyuk Lee, Seongjun Yun, Hyunjae Kim, Miyoung Ko, Jaewoo Kang

Published Date: 1 Oct 2018 Source: IEEE

Methodology: Open-domain QA systems are constructed as pipelines that include a retrieval system and

a reader model. We additionally built Paragraph Ranker that assists our QA pipeline for a better paragraph selection. For the retrieval system and the reader model, we used Document Retriever

and Document Reader. Paragraph Ranker- Answer Aggregation.

Pros and cons: By using Paragraph Ranker, the QA pipeline benefits from increased answer recall

from paragraphs to read, and filters irrelevant documents or paragraphs. With our simple Paragraph

Ranker, we achieve state-of-the-art performances on the four open-domain QA datasets with large

margins. As future works, we plan to further improve Paragraph Ranker based on the researches

on learning to rank.

1. **BERT: Pre-training of Deep Bidirectional Transformers for**

**Language Understanding**

Author/s: Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova

Published Date: 11 Oct 2018 Source: IEEE

Methodology: There are two steps in this framework: pre-training and fine-tuning. During

pre-training, the model is trained on unlabelled data over different pre-training tasks. For finetuning,

the BERT model is first initialized with the pre-trained parameters, and all of the parameters are fine-tuned using labelled data from the downstream tasks. Each downstream task has separate fine-tuned models, even though they are initialized with the same pre-trained parameters. Pre-training BERT- Fine-tuning BERT.

Pros and Cons: Recent empirical improvements due to transfer learning with language models have demonstrated that rich, unsupervised pre-training is an integral part of many language understanding systems. In particular, these results enable even low-resource tasks to benefit from deep unidirectional architectures.

1. **DocChat: An Information Retrieval Approach for Chatbot Engines**

**Using Unstructured Documents**

Author/s: Zhao Yan, Nan Duan, Junwei Bao, Peng Chen, Ming Zhou, Zhoujun Li, Jianshe Zhou

Published: 2016 Source: IEEE

Methodology: Considering a much simplified task, short text conversation (STC) in which the response R is a short text and only depends on the last user utterance Q. Previous methods for the STC task mostly rely on Q-R pairs and fall into two categories: Retrieval-based methods (e.g., Ji et al., 2014). This type of methods first retrieve the most possible hQˆ, Ri ˆ pair from a set of existing Q-R pairs, which best matches current utterance Q based on semantic matching models, then take Rˆ as the response R. One disadvantage of such a method is that, for many specific domains, collecting such QR pairs is intractable. Generation based methods. This type of methods usually uses an encoder-decoder framework which first encode Q as a vector representation, then feed this representation to decoder to generate response R. Similar to retrieval-based methods, such approaches also depend on existing Q-R pairs as training data. Like other language generation tasks, such as machine translation and paraphrasing, the fluency and naturality of machine generated text is another drawback. To overcome the issues mentioned above, we present a novel response retrieval approach, DocChat, to find responses based on unstructured documents. For each user utterance, instead of looking for the best Q-R pair or generating a word sequence based on language generation techniques, our method selects a sentence from given documents directly, by ranking all possible sentences based on features designed at different levels of granularity. On one hand, using documents rather than Q-R pairs greatly improve the adapt.

Pros and Cons: One shortcoming of the unsupervised topic model is that, the topic size is pre-defined, which might not reflect the truth on a specific corpus. Response retrieval cannot always

guarantee to return a candidate set that contains at least one suitable response, but response ranking will output the best possible candidate all the time. So, we have to decide which responses are confident enough to be output, and which are not.

1. **Optical Recognition of Braille Writing Using Standard Equipment**

Author/s: Jan Mennens, Luc van Tichelen, Guido Francois, and Jan J. Engelen

Published Date: Dec 1994 Source: IEEE

Methodology: Using the property that Braille characters are always positioned on a fixed matrix, it first tries to build a grid consisting of horizontal and vertical lines that run through all the dots, and then it checks if there is a dot present on each of the intersection points. The grid construction must be flexible because there are cases where it can be deformed or irregular. This depends greatly on the quality of the original.

Pros and Cons: Although Braille dots are placed on a fixed matrix, certain Braille production techniques cause this matrix to be irregular. - Deformations of the Braille Cell-Deformation of the Grid Where Braille Characters are Positioned.

The main reason for developing a system that can read Braille is to preserve and multiply large volumes of manually crafted books. Many books on mathematics or music are very difficult, even for a skilled copyist, to retype due to the special rules that apply in Braille.

1. **Teaching Machines to Read and Comprehend**

Author/s: Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, Lei Yu, and Phil Blunsom

Published Date: 10 Jun 2015 Source: IEEE

Methodology: Distributed/neural models allows us to learn shallow features for our

classifiers, capturing simple correlations between inputs. Recurrent neural networks provide a very practical tool for sentence embedding. Viewing translation as encoding and decoding sentences.

Pros and Cons: Twenty years ago log-linear models allowed greater freedom to model correlations than simple multinomial parametrisations, but imposed the need for feature engineering. This project eradicates the problem.

1. **DRISHTI—A Gesture Controlled Text to Braille**

**Converter**

Author/s: Vineeth Kartha, Dheeraj S. Nair, Sreekant S., Pranoy P. and Dr. P. Jayaprakash

Published Date: 28 January 2013 Source: IEEE

Methodology: The device consists of a microcontroller which is the heart of the device. The conversion of characters take place in this device, at present we have started the work with Atmega 8 chip as it was available to us but the chip had to be changed to Atmega 328 due to memory constraints of Atmega 8. The code of Atmega 8 is compatible with Atmega 328, so switching the chip will not hinder the progress made so far. The tactile display is made of six solenoids that represent the Braille characters, the device will be having only a single Braille cell. Atmega328- Touchpad The microcontroller code using the Arduino platform. Software on the computer using processing platform, poppler utils and unoconv

Pros and Cons: suggestion made by a person who tested this device was that by including upto 5 braille cells the speed problem can be solved. Also since the device we made was a prototype it had imperfections in the leveling of the braille cell. use of piezoelectric Braille cell will further improve speed of the device and also reduce the power consumption. The present device is noisier due to the solenoids using piezo-electric braille cells will reduce the noise too. Another development possible is to find suitable batteries that can be used to power up the device so that it can be used even without AC mains. Integrating USB drives in place of Micro SD cards is also a possible expansion on the hardware side. On the software side using optical character recognizers opens a new arena for the visually challenged to learn and read. The Braille displays are superior to text to speech synthesizers as it promotes Braille literacy.

1. **Development of a Text to Braille Interpreter for Printed**

**Documents through Optical Image Processing**

Author/s: Joshua L. Dela Cruz, Jonaida Angela D. Ebreo, Reniel Allan John P. Inovejas,

Angelica Romaine C. Medrano, and Argel A. Bandala

Published Date: 3 Dec. 2017 Source: IEEE

Methodology: Haptic System, OCR, Image Processing, Text

Conversion. - the camera will capture an image, in tiff format, which will be analyzed through an optical image processing integrated in a GUI. After analysis, a universal asynchronous receiver/transmitter (UART) device will interface the communication between the computer and the

microcontroller. The microcontroller then sends the data to the refreshable braille cells.

*Camera- Optical Image Processing- USB to UART Device- Microcontroller- Braille Cells*

Pros and cons: the solution presented by this study is better than the similar and existing studies because the development of this system is definitely more modernized or smart compared to the existing ones. This system will give the visually-impaired people the chance to read any printed reading paraphernalia. The system is able to achieve an overall system speed

of 1 word in 2 seconds, and an overall system reliability of 95.68%. It is easy to use but is not stand-alone. The system is efficient both in character recognition and in actuation of braille cells, making it a good alternative to braille printed books.