Automatic Speech Recognition and Machine Learning for Robotic Arm in Surgery

J Ruby[1], Susan Diana[2], Yanmin Yuan[3], William Harry[4], J Tisa[5], J Nedumaan[6], Yang Yung[7], J Lepika[8], Thomas Binford[9], P S Jagadeesh Kumar[10] and Wenli Hu[11]

[1,2]*Department of Medical Sciences, University of Oxford, United Kingdom*

[3]*Department of Bioengineering, Harvard University, United States*

[4]*Center for Biomedical Imaging, Stanford University, United States*

[5]*Stanford Center for Biomedical Informatics Research, Stanford University, United States*

[6,9,1]0*Department of Computer Science, Stanford University, United States*

[7,11]*Biomedical Engineering Research Center, Nanyang Technological University, Singapore*

[8]*Department of Computer Science, Harvard University, United States*

Submission: December 20, 2019; **Published:** March 05, 2020

*Corresponding author: J Ruby, Department of Medical Sciences, University of Oxford, United Kingdom

Abstract

This article liberalizes the current machine learning rehearses as utilized in the emerging edge and as noteworthy to speech recognition approaches on present-day surgical robots. The desire is to advance the development of medical robots among the machine learning and speech recognition liberal that has spread out from the point of view of health care services in social protection. The propelled subordinates are sifted through the central machine learning models that are pattern setting and have a show for center and progress that are steady to robotic arm and speech recognition practices. Machine learning is displayed in the comprehension of speech recognition components and its influence in biomedical robots for surgeries. Topical advances of machine learning and intelligent algorithms, further accentuations on their vast hugeness in the improvement of speech recognition in medical surgical applications.

Keywords: Automatic speech recognition; Surgical robots; Robotic arm; Machine learning

Abbreviations: ASR: Automatic Speech Recognition; ML: Machine Learning; WER: Word Error Rate

Introduction

The significant desire of this article is to propose knowledge from various locations while systematizing numerous Automatic Speech Recognition (ASR) procedures into a dug in Machine Learning (ML) technique. Further unequivocally, this paper offers a general thought of wide ASR techniques by making various methods for arranging and grouping the incessant ML paradigms, worked by their learning style. The learning habits upon the characterization of the learning techniques that are alluded to the key qualities of the ML algorithms, similar to the component of the calculation's info or yield, the choice capacity practiced to set up the order or recognizable proof of yield, and the hardship used in preparing the understandings [1]. While tangling on the key recognizing factors related with the different gatherings of the ML algorithms,

remarkable thought is remunerated to the related figure urbanized in ASR investigate. In its wide degree, the attempt of ML is to develop programmed plans capable of streamlining from heretofore watched examples, and it understands so by making or learning pragmatic dependence during the abstract info and yield domains. ASR, which is foreseen to interpret the sound data in speech insights into its principal phonetic example, traditionally as recognition arrangement, is hence basically a ML difficulty, i.e., determined cases of impacts as the ceaseless sound trademark arrangement or maybe solid waves and yields as the apparent esteemed mark arrangement (presentation, voice, or articulation), the goal are to imagine the most recent creation arrangement from a unique info string. This prescience task is regularly named as order when the consecutive segment limitations of the yield names are comprehended as known [2]. If not, the prescience task is called ID. For example, phonetic grouping and phonetic recognizable proof are two various tasks: the prior with the telephone limitations given in both preparing and testing information, while the last involves no such outskirts information

and are consequently further unpredictable. Correspondingly, confined phrase "recognizable proof" is a run of the mill scientific categorization perspective in ML, aside from with an irregular viewpoint in the info space attributable to the impulsive length of the speech commitment. Besides ceaseless, speech recognition is an excellent sort of organized ML tribulations [3].

The use of robots in medical surgical applications isn't new. Robots starting at now help in brain and spinal medicinal technique, with models, for instance, renaissance empowering masters to fix spines with 99 percent accuracy (9 percent higher than conventional strategies). The well-known da Vinci cautious system (where master's hand developments are changed over into tinier, progressively definite mechanical improvements) is right now used over a wide extent of dealings, from prostate harmful development treatment to performing heart valve restorative method [4]. In the US, a robot called Watson helps ends and creates the administrators' arrangements for oncology patients by mixing information from an enormous number of reports, getting records, clinical starters and bulletins. Meanwhile, Woebot, the world's first mechanical specialist, has more than 2,000,000 dialogs consistently [5]. Regardless of the way that experts at the Children's National Medical Center in Washington have starting late developed a surgical robot (called STAR) which can suture fragile tissue; experts says we're still a very long time from having the prior referenced C-3PO-style robots in our theaters. The analysis is that still have far to go before a patent insignificant exertion with enough skill and affectability in robots are attempted to play out the kind of work discussed.

Materials and Methods

From a commonsense view, ASR is the change strategy from the acoustic data course of action of speech into a word gathering. From the specific point of view on ML, this change method of ASR requires different sub-structures including the usage of discrete time stamps, as often as possible called edges, to depict the speech waveform data or acoustic features, and the use of obvious imprints to record the acoustic data gathering [6]. The chief issues in ASR lie in the possibility of such names and data. It is basic to clearly appreciate the stand-out properties of ASR, to the extent both data and yield names, as a central motivation to relate the ASR and ML investigate domains and to esteem their spread. From the yield viewpoint, ASR yields sentences that involve a variable number of words. Thusly, at any rate on a crucial level, the amount of possible classes for the portrayal is huge to the point that it is in every way that really matters hard to create ML models for complete sentences without the usage of structure [7]. From the information viewpoint, the acoustic data are other than a course of action with a variable length, and generally, the length of data input is immeasurably not equivalent to that of imprint yield, offering climb to the unprecedented issue of division or game plan that the "static" gathering issues in ML don't encounter

[8]. Merging the data and yield viewpoints, the chief issue is communicated as a composed course of action request task, where a progression of acoustic data is used to understand a gathering of the semantic units, for instance, words. It is significant that the gathering structure in the yield of ASR is usually more astounding than most by far of request issues in ML where the yield is a fixed, restricted plan of classes [9]. Further, when sub-word units and setting dependence are familiar with create sorted out models for ASR, basically more imperative multifaceted nature can develop than the reasonable word method yield in ASR.

Even more fascinating and stand-out issue in ASR, in any case, is on the information side, explicitly, the variable-length acoustic-incorporate gathering. The extraordinary typical for speech as the acoustic commitment to ML counts makes it at times more problematic antiquity for the assessment than other [10]. In that limit, in the ordinary ML composing, there has usually been less highlight on speech and related transient plans than on various standards and models. The exceptional typical for speech lies fundamentally in its transient estimation; explicitly, in the huge variance of talk related with the adaptability of this common estimation. Therefore, paying little respect to whether two yield word progressions are indistinct, the speech data routinely have specific lengths, e.g., particular data tests from a comparable sentence as a general rule contain different data dimensionality depending upon how the talk sounds are conveyed [11]. Further, the discriminative signs among separated speech classes are frequently appropriated over a reasonably long common range, which normally crosses neighboring talk units. Other exceptional pieces of talk incorporate class-subordinate acoustic sign. These signs are normally conveyed over different time crosses that would benefit by different lengths of assessment windows in talk examination and feature extraction. Finally, perceived from other gathering issues for the most part inspected in ML, the ASR issue is a remarkable class of sorted out model affirmation where the apparent models, are embedded in the general transient course of action plan. Standard perspective places that speech is a one dimensional common upright instead of image and video as higher dimensional signs [5]. This view is summed up and doesn't get the exemplification and difficulties of the ASR issue. Dialogue is best observed as a two-dimensional standard, where the spatial and momentary estimations have inconceivably extraordinary characteristics, instead of images where the two spatial estimations will when all is said in done have relative properties. The spatial estimation in dialogue relates to the repeat proliferation and related changes, getting different variety types including basically those rising out of circumstances, speakers, accentuate, talking style and rate. The last sort actuates relations during spatial and common estimations, and the earth factors incorporate enhancer characteristics, speech transmission, encompassing disturbance, and room reverberation. The short-lived estimation in speech and its explicit association with the spatial or repeat region properties of dialogue sets up one of the surprising troubles for ASR [3,4].

A segment of the advanced generative models related with the generative learning perspective of ML, where Bayesian philosophies are used to give transient impediments as prior data about the human speech age process. Using the possibility of the adversity work similarly as the decision limit, the noteworthy ML perfect models are isolated into generative and discriminative learning orders. Dependent upon what kind of getting ready data is open for learning; then again ML perfect models are masterminded into coordinated, semi-oversaw, single, and dynamic learning classes [9,10]. Right when difference among source and target apportionments raises, a more commonplace condition in ASR than various zones of ML applications, the ML perfect models are orchestrated into single task; perform different assignments, and flexible learning [11]. Finally, using the trait of data depiction has meager learning and significant learning models, both later progressions in Machine Learning and Automatic Speech Recognition.

Results and Discussion

Figure 1: Speech Assisted Surgical Robotic Arm in Hip Replacement Surgery.

Figure 2: A Typical Surgical Robotic Arm Featuring Automatic Speech Recognition.

Robotic medical surgical procedure is a training to acknowledge medical procedure by methods for little instruments committed to a mechanical arm. The specialist controls the mechanical arm with a computer. The new research bearing expressed in this paper is to control the robotic arm through speech processing handling not at all like computer programming. For often starting at now, the mechanical robotic arm developments have been constrained by the specialist hand development. Over the most recent couple of years broad research had been led by Oxford Medical Research Center toward this path. Table 1, (Figure 1 & 2) demonstrates machine learning based speech recognizing mechanical robotic arm in medical surgical procedures. The outcomes unmistakably show the amazingness of machine learning adapting automatic speech recognition in medical surgical procedures. Word error

rate (WER) is a typical measure of the enactment of automatic speech recognition on machine interpretation framework. However, to the disclosure, machine learning and customary ASR techniques gave significantly progressively better exactness. Machine learning methods like Hidden Markov Models provided higher word rate error compared to Dynamic Time Warping and Conditional Random Fields.

Table 1: Performance of Machine Learning Based Speech Assisted Surgical Robotic Arm.

Machine Learning Technique	Merit	Word Error Rate (%)
Hidden Markov Models	Strong Statistical Foundation	60.8
Dynamic Time Warping	Feasible Word Range	54.3
Conditional Random Fields	Flexible Feature Selection	50.3

Conclusion

The progressing robotic medical surgical research vamped the excitement for speech processing systems with AI, distinctive measurement depictions from both ASR and ML group utilizing more predominant streamlining practices than in the past is an instance of the assessment moving towards this advance. To obtain complete natural speech processing robotic-assisted surgery is an endeavor that will require associated ML schemes inside and possibly past the perfect models. To append, automatic speech recognition based mechanical medical procedure with no uncertainty is going to manage robotic medical surgical field in the decades to come. However, ML based speech recognition gave impressive word precision, the authors' future research point of view is to distinguish a devoted ML algorithm for strong automatic speech recognition based robotic surgical procedure to quick track both precisely and financially feasible innovation.
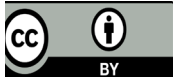
Acknowledgement

References

1. Moustris GP, Hiridis SC, Deliparaschos KM, Konstantinidis KM (2011) Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature. International Journal of Medical Robotics and Computer Assisted Surgery 7(4): 375-392.

2. Challacombe B, Stoianovici D (2009) The basic science of robotic surgery. Urologic Robotic Surgery in Clinical Practice: 1-23.

3. Reynolds DA (2002) An overview of automatic speaker recognition technology. In Proc. IEEE Int. Conf. Acoust, Speech Signal Process 4: 4072-4075.

4. Morgan N (2012) Deep and wide: Multiple layers in automatic speech recognition. IEEE Trans. Audio Speech Lang Process 20(1): 7-13.

5. Jagadeesh KPS, Yanmin Y, Ruby J, Yang Y (2018) Deep Robotics in Computational Neuroscience and Human Brain Behavior. The World Congress on Engineering and Computer Science (WCECS'18) Proceedings of International Conference on Intelligent Automation and Robotics (ICIAR'18) San Francisco, USA 23-25: 66-71.

6. Jagadeesh KPS, Yanmin Y, Yang Y, Mingmin P, Wenli H (2018) Robotic Simulation of Human Brain Using Convolutional Deep Belief Networks. Int J Intelligent Machines and Robotics 1(2): 180-191.

7. Gomes P (2011) Surgical robotics: reviewing the past, analysing the present, imagining the future. Robotics and Computer-Integrated Manufacturing 27(2): 261-266.

8. Jagadeesh KPS, Mingmin P, Xianpei L, Yanmin Y (2017) Modern Machine Learning Conniptions for Automatic Speech Recognition. International Journal of Information and Communication Technology 44(10): 150-160.

9. Jagadeesh KPS (2018) Deep Learning Based Cognitive Self-Supervised Robots in Medical Engineering. 6th World Convention on Robots, Autonomous Vehicles and Deep Learning, Smart Robotics Congress, September 10-11.

10. Jagadeesh KPS, Xianpei L, Thomas B (2019) Deep Robotics and Medical Applications. Baker & Taylor Publisher Services, North Carolina, United States. Applications-Book.

11. Kenngott H, Fischer L, Nickel F, Rom J, Rassweiler J, et al. (2012) Status of robotic assistance: a less traumatic and more accurate minimally invasive surgery? Langenbeck's Archives of Surgery 397(3): 333-341.