Asian Journal of Research in Computer Science

11(1): 1-7, 2021; Article no.AJRCOS.71452
ISSN: 2581-8260

Sign Translation Via Natural Language Processing

Shadman A. Khan[1*], Zulfikar Ali Ansari[1], Riya Singh[1], Mohit Singh Rawat[1], Fiza Zafar Khan[1] and Shubham Kumar Yadav[1]

[1]Department of Computer Science, Babu Banarasi Das National Institute of Technology and Management – 054, (of Dr. APJ AKTU), Lucknow, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJRCOS/2021/v11i130251
Editor(s):
(1) Dr. Xiao-Guang Lyu, Jiangsu Ocean University, China.
Reviewers:
(1) Ileana Hamburg, Westphalian University, Germany.
(2) Ajay Prasad, University of Petroleum and Energy Studies, India.
Complete Peer review History: https://www.sdiarticle4.com/review-history/71452

Original Research Article

Received 01 June 2021
Accepted 06 August 2021
Published 07 August 2021

ABSTRACT

Artificial Intelligence (AI) technologies are new technologies with new complicated features emerging quickly. Technology adoption has been beneficial for many general models. The models help in train the voice user-interface assistance (Alexa, Cortona, Siri). Voice assistants are easy to use, and thus millions of devices incorporate them in households nowadays. The primary purpose of the sign language translator prototype is to reduce interaction barriers between deaf and mute. To overcome this problem, we have proposed a prototype. It is named sign language translator with Sign Recognition Intelligence which takes the user input in sign language and processes it, and returns the output in voice out load to the end-user.

Keywords: Voice user interface; natural language processing; artificial intelligence; Tensor flow; k-nearest neighbors algorithm.

1. INTRODUCTION

The system aims to get the deaf and dumb people more involved in communicating. The camera or webcam would be helpful in terms of conversation with a dumb person and capturing the signs. Also, they could use it to recognize and convert sign language gestures into plain

**Corresponding author: E-mail: contactshadman@gmail.com;*

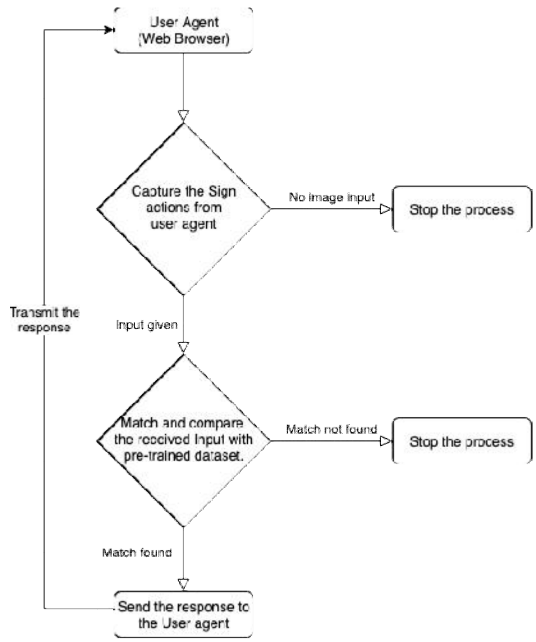

Fig. 1. Workflow for translating the input sign

text, i.e. English and then to their original language. Our main objective is to design a simple solution for most people with deaf and dumb. There are also many methods to solve sign language like "Kinect" to solve the language and get the input and work on them. Nevertheless, "Kinect" is very much complicated to decipher and understand. Our approach is to provide our users with a simple and better way to complete the task with no bugs, and we have used standard and easily available libraries in our system [1].

Now, a day's Amazon has launched new translation features that allow users to speak in two different languages while communicating with each other, with Alexa acting as an interpreter and translating both sides of the conversation. Once the session has commenced (begin), the customer can speak phrases and sentences in other languages. Alexa will automatically identify which language has been spoken and translate each side of the conversation.

It works with eight language pairs – English, Spanish, French, German, Italian, Brazilian, Hindi, or Portuguese [2,3].

In 2011, 1.3 million people were diagnosed with "hearing impairment." However, India's National Association of the Deaf has reported approximately 18 million people, i.e. 1 per cent of the population [4,5].

2. RAPID GROWTH RATE

Artificial Intelligence (AI) has shown significant progress in recent years, and its potential is growing. AI technologies are one of the new technologies with new complicated features emerging at a fast pace. An application area of AI is Natural Language Processing (NLP). Voice-activated personal assistants (VAPAs)--like Amazon Echo or Apple Siri--offer considerable promise to individuals who are blind due to the widespread adoption of these non-visual interaction platforms. Technology adoption has been a part of the study for many years, and there are many general models in the literature describing it. Nowadays, search engines are processing Open Data to know what kind of data there is to get there would be of help. This paper presents a voice assistant which uses Open Data as its knowledge source [6,7].

2.1 Impact on Society

Voice user interface (VUI) has been very complicated in terms of implementation for developers and consumers. Consumers feel high pressure before interacting or performing any critical task (payments and reviews). However, few tech giants like Amazon, Microsoft and

Apple, have made it easy to understand and configure. Nevertheless, voice technology benefits those with disabilities such as visual and hearing impairment [8]. Because the devices with the ability to interact with the user's voice can be connected to other utilities over a similar network, access can be given to people who may not operate a remote control, a light switch, or see the TV listings on the screen [9-16].

3. METHODOLOGY

The main objective is to read and understand the captured images from the webcam and camera and extract the right meaning. The prototype requires an established local network system. The runtime environment for the prototype requires Node.js, a JavaScript runtime built on Google Chrome's V8 JavaScript engine, installed on a local machine. It uses several web APIs (Application programming interface) available on the web browser or user agent, which has exposed the robust environment to achieve the proper functionality for converting Text to Speech or vice versa [17].

The other objective of this prototype is to create a machine that can be teachable without many complexities. The prototype practices a kNN (k-Nearest-Neighbours) approach which is easily understandable that it technically does not perform any "learning" at all. Despite, it takes an input image (from the webcam or external camera) and classifies it by finding the label of coaching examples closest to the present input image employing a similarity function or distance metric [18]. However, before feeding the captured image data set to kNN, the image is first passed through a deep neural network called Squeeze Net. Squeeze Net is a convolutional neural network with 18 layers and applies design strategies to reduce the number of parameters, notably with fire modules that "squeeze" parameters using 1x1 convolutions [19]. A Fire module comprises a squeeze convolution layer (which has only 1x1 filters), feeding into an expand layer with a mix of 1x1 and 3x3 convolution filters. The output from the penultimate layer of this network is then fed into the kNN, which allows you to coach your classes. The advantage of doing it in this manner, rather than directly feeding raw pixel values from the webcam into the kNN, is that we will use the high-level abstractions that Squeeze Net has already learned, thus training a better classifier [20-23].

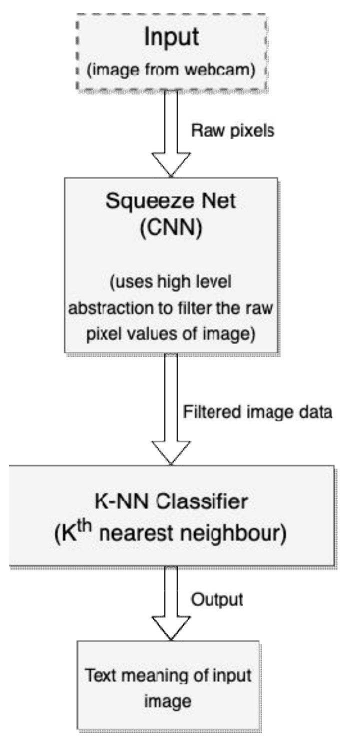

Fig. 2. Flow chart of image processing

A highly immerse computer-primarily based approach for performing a command via a voice consumer interface on a subset of objects (in this case, image data sets). The subset is selected from fixed items, each having an object type. At least one taggable field (key) is associated with the object type and has a corresponding value—the set of objects stored in the system's memory. An utterance is acquired from the person and consists of a command, an object type choice, a tag-gable field selection, and a price for the taggable discipline [24].

4. RESULTS AND DISCUSSION

The methodology used uses the k-nearest neighbors (KNN) algorithm to identify the sign language taken from the user. The captured raw pixels from the webcam go through a deep neural network (Squeeze Net) scan or assume it as a filter, which ultimately returns selective or matched data of raw pixels, then it is passed into the KNN Classifier.

Training a classifier was one of the crucial steps since everything is running at the program's runtime.

The application had successfully identified the captured images from the webcam with their respective semiotic meaningful texts.

There could be several other ways to approach this problem, which may serve as beneficial starting points.

Google's Tensorflow.js has released a PoseNet can be used to estimate either a single pose or multiple poses and using this could be an exciting approach. From the machine's standpoint, tracking the wrist, elbow, and shoulder position should be competent to predict most words. Finger positions tend to matter most when spelling something out [25,26].

Using the CNN-based approach could improve the detection time and help make the model more resistant to translational invariances, which could occur from different users of different signs style. This approach would also include saving a model or load a pre-trained Keras model, well documented. Using this approach would remove the need for training the system every time the user restarts the browser [27-31].

5. RELATED WORK

Creating successful sign language processing machines needs an understanding of Deaf culture to create systems that align with user needs and desires, and of sign languages to build systems that account for their complex linguistic aspects [32-34]. Here, we summarize this background, and we also discuss existing reviews of sign language processing, which do not take a comprehensive view of the problem.

5.1 Deaf Culture and ASL

Most sign language users are a cultural minority with no common language or practice. Many people read deafness is not disabled as cultural identity with many benefits [35,36].

Sign language is completely in their hands, the movement of the eyebrows, mouth, head, shoulders, and eyes. For example, ASL upper brows are most likely an open question, while frowned brows indicate a yes/no question. Signs can also be entered by placing actions in the mouth, for example, chief executive characters, CUPS different mouth positions can be set using the cup size. Sign language recognition software must be carefully not found by hand components [17,37].

There is a wide variation in sign language performance based on ethnic, racial, geographical region age, gender, educational level, language, knowledge, hearing status, etc., in spoken language in different social and geographical groups and the use of different kinds of people (for example, Black ASL differs from a sign language dialect). Unlike spoken language, sign language makes use of different topics [38-43]. Most deaf children are born from hearing parents who may not know non-sign language when the child was born [44]. Thus, in the deafest sign language users, learning foreign languages followed from childhood or into adulthood tends to cause a decrease in flow. Sign language processing software there are accurate models and identifying richness requires increasing the volume and diversity and training of information. It is difficult to determine whether it is a sign of the volume of vocabulary. Available ASL-to-English dictionary contains 5-10 thousand characters. But they are representative of the real world: no classifiers, images, or other means, a feature of the signal input to add an adjective, adverb, and nuanced meaning to the word [36,45].

6. CONCLUSION

Various methods could have solved the problem statement. The methodology used was relatively straightforward since the focus of this prototype is addressing the problem of people with disabilities and help them efficiently and effectively. Any of the world's most prominent technology groups could have developed this feature. Even Amazon has also developed a way to communicate with Echo devices by pressing buttons on them.

On the Amazon website, we can read that "With natural language understanding (NLU), computers can deduce and extract meaning what a speaker means, and not just the words they say. It facilitates voice technology like Alexa to indicate that the user would probably ask for a local weather forecast when the user asks, "Alexa, what is it like outside."

The aim of building this prototype was not to solve the entire sign language to text problem. Instead, it was to initiate a conversation around inclusive design, present machine learning in an approachable light, and inspire people to explore this problem space — something I hope this project achieved.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Graeme McLean, Kofi Osei-Frimpong. Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants; 2019.
2. Amrita S Tulshan, Sudhir Namdeorao Dhage. Survey on virtual assistant: Google assistant, siri, cortana, alexa; 2019.
3. Kawamura T, Ohsuga A. Flower voice: virTual assistant for open data; 2013.
4. Philippe Dreuw, Hermann Ney. Towards automatic sign language annotation for the elan tool. In Workshop Programme. 2008;50.
5. Sarah Ebling, John Glauert. Building a Swiss German sign language avatar with JASigning and evaluating it among the Deaf community. Universal Access in the Information Society. 2016;15(4):577–587.
   DOI: http://dx.doi.org/10.1007/s10209-015-0408-1
6. George Terzopoulos, Maya Satratzemi. Voice Assistants and Smart Speakers in Everyday Life and in Education; 2020.
7. Abdolrahmani Ali, Kuber Ravi, Branham Stacy. "Siri talks at you": An empirical investigation of voice-activated personal assistant (VAPA) usage by individuals who are blind. 2018;249-258.
   DOI: 10.1145/3234695.3236344.
8. Melvin Johnson. Google's multilingual neural machine translation system: Enabling zero-shot translation; 2016.
9. Elshafei M. Virtual personal assistant (VPA) for mobile users; 2002.
10. Samuel R Bowman, et al. Generating sentences from a continuous space, CoNLL; 2016.
11. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks; 2015.
12. Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, Li Fei-Fei. Large-scale video classification with convolutional neural networks; 2014.
13. Diksha Khurana, Aditya Koli, Kiran Khatter, Sukhdev Singh. Natural language processing: State of the art, current trends and challenges; 2017.
14. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep residual learning for image recognition; 2015.
15. Gao Huang, Zhuang Liu, Laurens van der Maaten. Densely connected convolutional networks; 2016.
16. Liao Q, Poggio T. Bridging the gaps between residual learning, recurrent neural networks and visual cortex; 2016.
17. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In CVPR. 2015;2.
18. Google: Google Assistant (developers.google.com/assistant/docs).
19. Classification of Scoliosis based on Multiclass SVM Model, Animita Das - Mtech Scholar, Dr.Tabitha Janumala - Assistant Professor, Department of Electrical and Instrumentation Engineering, 1RV College of Engineering, Bengaluru, India.
20. Lin M, Chen Q, Yan S. Network in network. In ICLR; 2014.

21. Netzer Y, Wang T, Coates A, Bissacco A, Wu B, Ng AY. Reading digits in natural images with unsupervised feature learning, 2011. In NIPS Workshop; 2011; 5.
22. Pezeshki M, Fan L, Brakel P, Courville A, Bengio Y. Deconstructing the ladder network architecture. In ICML; 2016.
23. Pleiss G, Chen D, Huang G, Li T, van der Maaten L, Weinberger KQ. Memory-efficient implementation of densenets; 2017.
24. Purington A, Taft JG, Sannon S, Bazarova NN, Taylor SH. Alexa is my new BFF: Social roles, user satisfaction, and personification of the amazon echo; 2017.
25. Romero A, Ballas N, Kahou SE, Chassang A, Gatta C, Bengio Y. Fitnets: Hints for thin deep nets. In ICLR; 2015.
26. Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus Piater, Hermann Ney. RWTH-PHOENIX-Weather: A large vocabulary sign language recognition and translation corpus. in international conference on language resources and evaluation. Istanbul, Turkey. 2012;3785–3789.
27. Rasmus A, Berglund M, Honkala M, Valpola H, Raiko T. Semi-supervised learning with ladder networks. In NIPS. 2015;3.
28. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge; 2014.
29. Sermanet P, Chintala S, LeCun Y. Convolutional neural networks applied to house numbers digit classification; 2012.
30. Marco Tulio Ribeiro. Beyond accuracy: Behavioral testing of NLP models with checklist. ACL; 2020.
31. Maksym Davydov, Olga Lozynska. Information system for translation into Ukrainian sign language on mobile devices. In 2017 12th International Scientifc and Technical Conference on Computer Sciences and Information Technologies (CSIT). 2017;1(IEEE): 48–51.
32. Paul Dudis. Depiction of events in ASL: Conceptual integration of temporal components; 2004.
33. Richard Clark Eckert, Amy June Rowley. Audism: A Theory and practice of audiocentric privilege. Humanity & Society. 2013;37(2):101–130.
34. Ralph Elliott, John RW Glauert, JR Kennaway, Ian Marshall. The development of language processing support for the ViSiCAST project. In ASSETS. 2000;4th.
35. Gong L. Intelligent virtual assistant.
36. López G, Quesada L, Guerrero LA. Alexa vs. Siri vs. cortana vs. google assistant: A comparison of speech-based natural user interfaces; 2017.
37. Ralph Elliott, John RW Glauert, JR Kennaway, Ian Marshall, Eva Safar. Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. Universal Access in the Information Society. 2008;6(4):375–391.
38. Michael Erard. Why Sign-language gloves don't help deaf people. The Atlantic 9; 2017.

    Available:https://www.theatlantic.com/technology/archive/2017/ 11/why-sign-language-gloves-dont-help-deaf-people/ 545441/
39. Fels SS, Hinton GE. Glove-talk: A neural network interface between a data-glove and a speech synthesizer. IEEE Transactions on Neural Networks. 1993;4(1):2–8.

    DOI: http://dx.doi.org/10.1109/72.182690
40. Jens Forster, Christian Oberdörfer, Oscar Koller, Hermann Ney. Modality combination techniques for continuous sign language recognition. In Iberian Conference on Pattern Recognition and Image Analysis (Lecture Notes in Computer Science 7887). Springer, Madeira, Portugal. 2013;89–99.
41. Jens Forster, Christoph Schmidt, Oscar Koller, Martin Bellgardt, aHermann Ney. Extensions of the sign language recognition and translation corpus RWTH-PHOENIX-weather. In International Conference on Language Resources and Evaluation. Reykjavik, Island. 2014;1911–1916.
42. Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumeé III, Kate Crawford. Datasheets for datasets. arXiv preprint arXiv:1803.09010; 2018.
43. Ann E Geers, Christine M Mitchell, Andrea Warner-Czyz, Nae-Yuh Wang, Laurie S

Eisenberg. CDaCI investigative team, and others. Early sign language exposure and cochlear implantation benefts. Pediatrics. 2017;140:1.

44. Kepuska V, Bohouta G. Next generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home); 2018.

45. Terzopoulos George, Satratzemi Maya. Voice assistants and artificial intelligence in education. 2019;1-6.
DOI: 10.1145/3351556.3351588

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle4.com/review-history/71452