


RESEARCH

Open Access



Deep learning-based question answering system for intelligent humanoid robot

Widodo Budiharto^{1*} , Vincent Andreas¹ and Alexander Agung Santoso Gunawan²

*Correspondence:
wbudiharto@binus.edu
¹ Computer Science
Department, School
of Computer Science,
Bina Nusantara University,
Jakarta 11480, Indonesia
Full list of author information
is available at the end of the
article

Abstract

Background: The development of Intelligent Humanoid Robot focuses on question answering systems that can interact with people is very limited. In this research, we would like to propose an Intelligent Humanoid Robot with the self-learning capability for accepting and giving responses from people based on Deep Learning and Big Data knowledge base. This kind of robot can be used widely in hotels, universities, and public services. The Humanoid Robot should consider the style of questions and conclude the answer through conversation between robot and user. In our scenario, the robot will detect the user's face and accept commands from the user to do an action.

Findings: The question from the user will be processed using deep learning, and the result will be compared to the knowledge base on the system. We proposed our Deep Learning approach, based on Recurrent Neural Network (RNN) encoder, Convolution Neural Network (CNN) encoder, with Bidirectional Attention Flow (BiDAF).

Conclusions: Our evaluation indicates that using RNN based encoder with BiDAF gives a higher score, than CNN encoder with the BiDAF. Based on our experiment, our model get 82.43% F1 score and the RNN based encoder will give a higher EM/F1 score than using the CNN encoder.

Keywords: Humanoid Robot, NLP, Big data, Deep learning

Introduction

Robot learning is a research field at the intersection of machine learning and robotics. It studies techniques to acquire novel skills or adapt to the environment through learning algorithms. A novel approach proposed by Perera V et al. [1] enables a mobile service robot to understand questions about the history of tasks it has executed. They frame the problem of understanding such questions as grounding an input sentence to a query that can be executed on the logs recorded by the robot during its runs, by defining a query as an operation followed by a set of filters. Robotician Angelo Cangelosi of the University of Plymouth in England and Linda B. Smith, a developmental psychologist at Indiana University Bloomington, have demonstrated how crucial the body is for procuring knowledge. "The shape of the robot's body, and the kinds of things it can do, influences the experiences it has and what it can learn from [2]. Learning from demonstration approaches focuses on the development of algorithms that are generic in their representation of the skills and in the way, they are generated. One of the most promising

approaches is those that encapsulate the dynamics of the movement into the encoding [3].

Several robots become our state of the arts of this research: Sophia, the social Humanoid Robot and first Citizen Robot in the world, developed by Hanson Robotics [4], Entertainment Robot, the close domain question answering robot, with answer base that constructed from the encyclopedia, developed by Oh et al. [5], Hospital Receptionist Robot, developed by Ahn et al. [6], that can express emotion, friendliness and can recognize the face. DialogFlow used by the Hospital Receptionist Robot for the conversation decision-making method. A Bilingual English and Japanese Talking Robot developed by Wilcock et al. [7], that can talk about various topics from Wikipedia and answer the question by processing the information from the article in Wikipedia. The language can be switched immediately by the Robot when the user asks it.

Our research goal is to make a question answering system using deep learning with self-learning capability in the Humanoid Robot. Previously we have tried to find a question answering system journal for Humanoid Robot, but we have not found another journal that discusses it properly. This journal is organized as follows: Sect. 1 is “[Introduction](#)”, Sect. 2 is “[Related works](#)”, Sect. 3 is “[Knowledge base systems and deep learning](#)”, Sect. 4 is “[Proposed method](#)”, Sect. 5 is “[Experimental results and discussion](#)”, Sect. 6 is “[Conclusion](#)”.

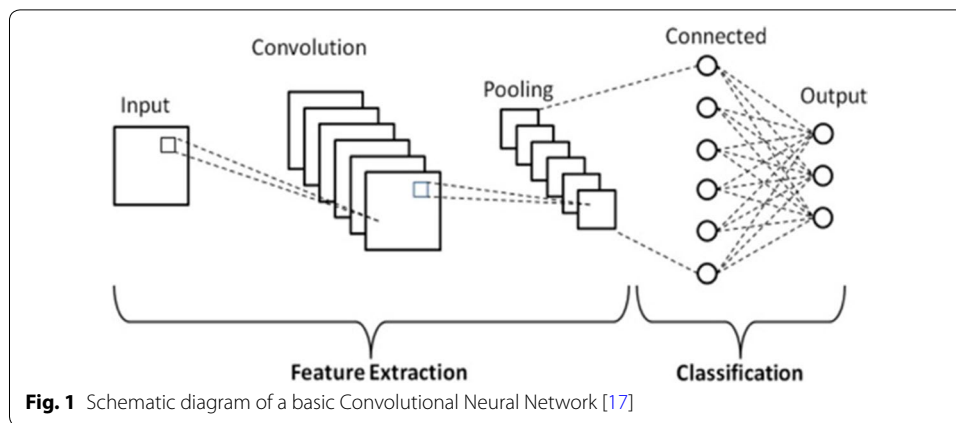
Related works

There have been previous efforts in exploring the question answering system. An approach proposed by Feng et al. [8] used two different approaches for answer selection. The first method uses the cosine similarity of the answer. The highest cosine value means the answer to the question. The second method using CNN. The result shows that the deep learning-based approach shows better results than another method.

The comparative study of CNN and RNN for NLP explored by Yin et al. [9] shows that RNN performs better than CNN in most of the NLP tasks. The model trained from scratch using the basic setup, and search optimal hyperparameters for each task and model separately to get the fair results.

Dependency Tree Recursive Neural Network proposed by Iyyer et al. [10] use for Factoid question answering over paragraphs. Dependency Tree Recursive Neural Network (DT-RNN) use to train the model with the dataset from quiz bowl tournaments. The model tested in two quiz categories: history and literature. The test shows that the model performance gets a higher score than the average human player in history question, but get lower score results in the literature question. From the experiments, they conclude that DT-RNN is an effective model for question answering, and can beat humans in some tests.

Another question answering approach proposed by Yin et al. [11], by using the Generative Question Answering (GEN QA) model to generate answers from factoid questions. The question will be transformed by Bidirectional RNN to representation form. The question will be compared with the knowledge base by calculating the relevance score using Bilinear Model or CNN-based Matching Model. In the last step, RNN will be used to generate the answer based on the relevance score result. The experiment result



shows that GEN QA based on the CNN model gets a higher score than using GEN QA based on the Bilinear model.

Another approach explored by Chen et al. [12] uses Wikipedia as the knowledge source to answer the question. They use Term Frequency–Inverse Document Frequency (TF-IDF) to rank the top 5 Wikipedia articles related to the question. The paragraph of articles and questions will be encoded using RNN. They predict the span by using the bilinear term to capture the similarity between paragraphs and questions.

Knowledge base systems and deep learning

Deep learning

Deep learning is a specific subset of Machine Learning, which is a specific subset of Artificial Intelligence. Computer vision and Natural Language Processing are examples of a task that Deep Learning has transformed into something realistic for robot applications. Using Deep Learning to classify and label images and text will be better than actual humans. Deep learning methods are proving very good at text classification, achieving state-of-the-art results on a suite of standard academic benchmark problems.

The advantage of Deep Neural Network (DNN), also referred to as deep learning, comes from its ability to extract high-level features from raw sensory data after using statistical learning over a large amount of data to obtain an effective representation of input space. This is different from earlier approaches that use hand-crafted features or rules designed by experts [13]. Two main DNN architectures that mainly used are CNN [14] and RNN [15].

CNN specialized kind of neural network for processing data that has grid-like topology. CNN has three types of layers: Convolution Layer, Pooling Layer, and Fully Connected Layer [16]. In the Convolution Layer, Convolution occurs by multiplication of two matrices. The illustration of CNN can be seen in Fig. 1.

The Pooling Layer is the next layer of CNN. Pooling Layer uses to reduce the input size and speed up the computation time. Two common functions used in the pooling operation are the Average Pooling Layer and Max Pooling Layer. Max Pooling Layer finds the maximum number in the $n \times n$ size window, usually, 2×2 , while the Average Pooling Layer finds the average number. In the fully connected layer, the result of the pooling layer will be used to classify or predict the result [16].

RNN excels in the sequence of data, like time series or sentence, but RNN has a problem in long-term dependency (the capability to remember information for a long period of time). Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) are the improvements of RNN. LSTM composed of a cell, an input gate, an output gate and a forget gate. Memory cells and gate units in LSTM learns to protect the constant error flow within the memory cell from perturbation by irrelevant inputs [18]. GRU consists of two gates, Reset Gate and Update Gate [19]. The Update Gate is the merging of Forget Gate and Input Gate in LSTM.

Knowledge base system

Big data is one of the most important part in deep learning. It will be used to provide data to the model. Big data can be used in many fields, in the NLP field, big data can be used to become the knowledge base by using Information Extraction. Information Extraction is the task of extracting structured information from unstructured text in an automatic fashion [20].

Knowing the entities (example: people, products, places, etc.) and class are important in constructing a knowledge base using textual big data. When entities can be identified, the extracted entities from the text can be canonicalized to registered entities using Named Entities Disambiguation (NED). The relationship between entities can be known by linking the entities [21].

Knowledge base engine proposed by Reshmi S and Balakrishnan K [22] integrates the database and Artificial Intelligence Markup Language (AIML). AIML used for analyzing missing format or information in the question that needed to answer the response in the chatbot. The Implementation example of analyzing textual big data is WISDOM X, which can discover the answer from the given question from around 4 billion web pages and can recommend related questions based on user search [23].

For the knowledge base, our Humanoid Robot can get the knowledge by 3 methods. The first method, Humanoid Robot automatically gets the data from the Big Data crawled website and extracts the information to become the knowledge in the knowledge base, The second method, the user manually inputs the knowledge, by saying the knowledge to the robot. The robot will listen and save knowledge to the knowledge base. The third method is inserting manually the knowledge by providing the knowledge base file to Humanoid Robot.

Proposed method

Our research focuses on question answering using the Deep Learning approach. The difference between our Humanoid Robot with the robot named Pepper Robot [24] is, Pepper covers Human Recognition, Object Recognition, and Speech Recognition using NAOqi [25]. NAOqi is the operating system for Humanoid Robot. In this research, we use Google API Speech Recognition in Python language to recognize voice [26]. To understand and find the answer, we use deep learning technology developed using Python. After finding the answer, we will use Google Text To Speech to speak the answer to the user.

In our previous work, we successfully proposed the face recognition and speech recognition system using stemming and tokenization for the Education Humanoid Robot. The

fun aspect is given as well because kids learn best when they are relaxed and focused. The Robot can give a good impact on student learning [27]. To improve the previous research, we want to make a Humanoid Robot with the ability of self-learning. The source of knowledge can be from text, web, or big data.

To achieve a complex behavior of the Humanoid Robot, it would be necessary to have inclusive and comprehensive repertoires of skills especially in response to the questions [28]. Our previous research is making Humanoid Robots for question-answering in the Indonesian language using Cosine Similarity [29].

Our algorithm for this Humanoid Robot is shown in Algorithm 1.

Algorithm 1. Self-Learning for Humanoid Robot.

```
function search_answer(user_question), do
    embed user_question sentence
    encode embed user_question sentence
    compare encoded user_question and encoded context to find the answer
    return answer

function add_to_knowledge_base(new_knowledge), do
    append new_knowledge to knowledge_base
    get append status
    return append status

if the camera detects the face, do
    greet

    while true, do
        listen to user sentence

        if the user asks a question, do
            listen to user's question
            get answer from search_answer(user_question)

            if answer not null, do
                say answer
            else, do
                say "not found"

        else if user want to add knowledge, do
            listen to knowledge from user
            add_to_knowledge_base(knowledge)

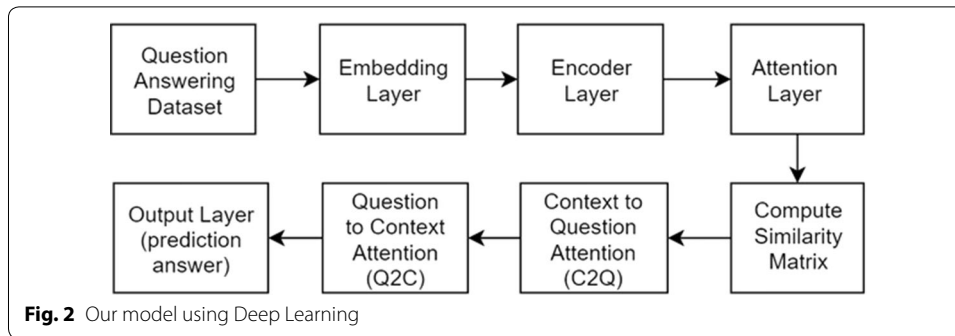
            if add to knowledge success, do
                say "successfully insert knowledge"

            else, do
                say "append knowledge process failed"

        else if user say goodbye, do
            say "bye-bye.. see you later"
            break
```

The model of Humanoid Robot using deep learning is shown in Fig. 2.

Briefly, the first step of the model is Question Answering (QA) Dataset, the QA Dataset will be converted to word embedding in Embedding Layer. The next step, Encoder Layer will be used to encode the dataset. Next, Attention Layer will be used to find the match between the hidden vector for dataset and hidden vector for a question, then compute Similarity Matrix. Next, we use Context to Question Attention



(C2Q), Question to Context Attention (Q2C), and combine them. The Output Layer will show the predicted answer.

For the training, QA Dataset needed to train the model. Stanford Question Answering Dataset (SQuAD) [30] will be used for the training dataset. This dataset based on Wikipedia articles with various topics, and contains 87.000 training questions answers (train dataset), and 10.000 development datasets (dev set). The answers' sentences in the SQuAD always part of the paragraph article.

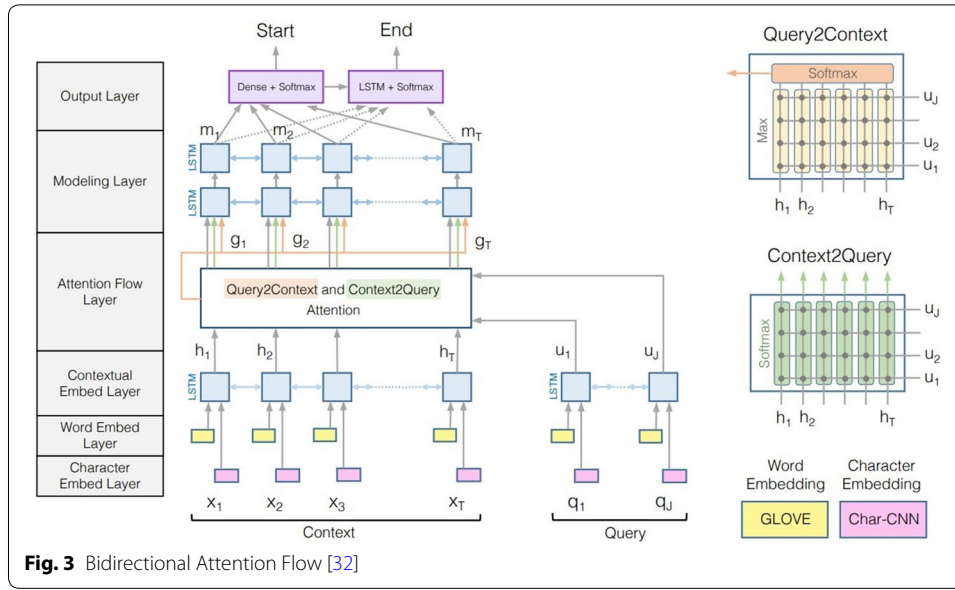
In the Embedding Layer, each word in the text will be converted to Word Embedding. Word Embedding is the representation of the word in the set of the vector. The word which has a similar meaning will have a similar representation of a vector. We use 100 dimensions of Global Vector (GloVe) [31] Word Embedding.

The next step is the Encoder Layer. The purpose of this step is to make a representation (encoding) for the dataset. We use two approaches, RNN (GRU/LSTM) based encoder and CNN based encoder. The output of the Encoder Layer is a hidden vector in the forward and backward direction. Then, we concatenate the hidden vector. Attention layer used to find the match between hidden vector for dataset and hidden vector for a question. We use Bidirectional Attention Flow (BiDAF) [32] as shown in Fig. 3.

The first step to use the BiDAF Attention Layer is the computing similarity matrix. $S \in \mathbb{R}^{N \times M}$, which contains a similarity score S_{ij} for each pair (c_i, q_j) of context and question hidden states. $S_{ij} = w^T \text{sim}[c_i; q_j; c_i \odot q_j] \in \mathbb{R}$. Here, $c_i \odot q_j$ is an elementwise product and $w \in \mathbb{R}^{6h}$ is a weight vector.

Next, we use Context-to-Question (C2Q) Attention. We take the row-wise softmax of S to obtain attention distributions α_j , which we use to take weighted sums of the question hidden states q_j , yielding C2Q attention outputs a_j . The next step is Question-to-Context (Q2C) Attention. For each context location $i \in \{1, \dots, N\}$, we take the maximum of the corresponding row of the similarity matrix, $m_i = \max_j S_{ij} \in \mathbb{R}$. Then we take the softmax over the resulting vector $m \in \mathbb{R}^N$ —this gives us an attention distribution $\beta \in \mathbb{R}^N$ over context locations. We then use β to take a weighted sum of the context hidden states c_i —this is the Q2C attention output c prime.

$$\beta = \text{softmax}(m) \in \mathbb{R}^N \quad (1)$$



$$c' = \sum_{i=1}^N \beta_i c_i \in R^{2h} \quad (2)$$

Finally, for each context position c_i , we combine the output from C2Q attention and Q2C attention as described in the equation below:

$$b_i = [c_i; a_i; c_i^\circ a_i; c_i^\circ c'] \in R^{8h} \forall i \in \{1, \dots, N\} \quad (3)$$

The final layer is a softmax output layer that use to decide the start and the end index for the answer span. The start and the end index used to determine which part of the paragraph is the prediction answer. The context of hidden states and the attention vector from the previous layer will be combined to create blended reps. These blended reps become the input to a fully connected layer which uses softmax to create a p_{start} vector with probability for start index and a p_{end} vector with probability for end index. We can look for start and end index that maximize $p_{start} * p_{end}$ [33, 34].

Experimental results and discussion

For the experiments, we test the model using RNN (GRU/LSTM) based Encoder and CNN based Encoder. We use the BiDAF attention layer, 150 hidden encoders, with 0.15 dropout and 33 batch size. We train using Nvidia GeForce RTX 2080 Super GPU with 8gb dedicated memory and 16gb shared memory, for 2 to 3 days.

During the training, the model will be evaluated by calculating the Exact Match Score (EM Score) and F1 score using the development dataset. After the model finished, we test them using the test dataset, 10% of the dataset used for the test dataset. The result shows in Table 1.

Table 1 Result of our experiment. It means that the RNN based Encoder gives better performance than using CNN based Encoder

Method	Dev (EM/F1)	Test (EM/F1)
RNN based Encoder	51.69/67.24	68.87/82.43
CNN based Encoder	46.35/61.34	53.99/69.55

EM measures the percentage of the prediction that matches one of the ground truth answers exactly. F1 measures the overlap between the prediction and ground truth answers which takes the maximum F1 over all of the ground truth answers.

For the big data analytical tools, we use Google BERT (Bidirectional Encoder Representations from Transformers) [35] as the comparison to find the best data analytical tools. Multi-layer Bidirectional Transformer Encoder [36] used by Google BERT as the model architecture. Next Sentence Prediction (NSP) also used to understand and predict the relationship between two sentences. From our test, the Google BERT model gets 90.9% in the F1 score, higher than our model, which only get 82.43% F1 score. Based on our comparison test, our model is quite good to do question answering system.

For models using RNN based Encoder, we get the optimum model at 93.000th iteration, while the model using the CNN based Encoder, we get the optimal result at 43.000th iteration. From two different approaches, RNN based Encoder gives better EM and F1 score results. The model could give the appropriate answers. The EM and F1 scores between dev and test have much better results, because we use 10% of training data, for testing data. So the answers produced have better quality in common. The proposed model successfully makes our Intelligent Humanoid Robot to accept questions and respond to the user with appropriate answers. Based on experiments we have done many times; our system has proven to be quite realistic and feasible to be used for real applications.

Conclusion

Our model is successfully obtained knowledge using big data technology and answer the questions from the user using deep learning. From our experiment using RNN and CNN as an encoder layer, we found that model with RNN based encoder and BiDAF attention layer, get higher EM and F1 scores, than the CNN encoder so the model can be used to handle question answering between Humanoid Robot and human. The RNN based encoder will give a higher EM/F1 score than using the CNN encoder.

For future development, we will implement the database to save knowledge, so the knowledge can store more data and manage easily. We will improve the algorithm to make better results in question answering and improve to handle unanswerable questions using the SQuAD 2.0 Dataset [37].

Abbreviations

AIML: Artificial Intelligence Markup Language; BERT: Bidirectional Encoder Representations from Transformers; BiDAF: Bidirectional Attention Flow; C2Q: Context-to-Question; CNN: Convolution Neural Network; DNN: Deep Neural Network; DT-RNN: Dependency Tree Recursive Neural Network; EM: Exact Match; GEN QA: Generative Question Answering; GloVe: Global Vector; GRU: Gated Recurrent Unit; LSTM: Long Short Term Memory; NED: Named Entities Disambiguation; NLP: Natural Language Processing; Q2C: Question-to-Context; QA: Question Answering; RNN: Recurrent Neural Network; SQuAD: Stanford Question Answering Dataset; TF-IDF: Term Frequency-Inverse Document Frequency.

Acknowledgements

This work is supported by Directorate General of Research and Development Strengthening, Indonesian Ministry of Research, Technology, and Higher Education, as a part of Penelitian Dasar Research Grant to Bina Nusantara University titled "Pemodelan Robot Humanoid dengan Kemampuan Self-Learning Berbahasa Indonesia" with contract number: 12/AKM/PNT/2019 and contract date: 27 March 2019.

Authors' contributions

All authors read and approved the final manuscript.

Funding

Penelitian Dasar Research Grant to Bina Nusantara University titled "Pemodelan Robot Humanoid dengan Kemampuan Self-Learning Berbahasa Indonesia" with contract number: 12/AKM/PNT/2019 and contract date: 27 March 2019.

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia.

² Mathematics Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia.

Received: 18 February 2020 Accepted: 14 August 2020

Published online: 16 September 2020

References

- Perera V, Veloso M. Learning to Understand Questions on the Task History of a Service Robot. In: IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Portugal. 2017; pp. 304–309.
- Kwon D. Self-Taught Robots, Scientific American; 2018 pp. 26–31.
- Billard A, Calinon S, Dillmann R, Schaal S. Robot programming by demonstration. In: Siciliano B, Khatib O, editors. Handbook of Robotics. Springer: USA; 2008. p. 1371–94.
- Retto J. Sophia, first citizen robot of the world. ResearchGate, URL: <https://www.researchgate.net> 2017.
- Oh HJ, Lee CH, Hwang YG, Jang MG, Park JG, Lee YK. A case study of edutainment robot: Applying voice question answering to intelligent robot. In: RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication 2007 Aug 26 (pp. 410–415). IEEE.
- Ahn HS, Yep W, Lim J, Ahn BK, Johanson DL, Hwang EJ, Lee MH, Broadbent E, MacDonald BA. Hospital Receptionist Robot v2: Design for Enhancing Verbal Interaction with Social Skills. In: 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) 2019 Oct 14 (pp. 1–6). IEEE.
- Wilcock G, Jokinen K, Yamamoto S. What topic do you want to hear about?: A bilingual talking robot using English and Japanese Wikipedias. In: The 26th International Conference on Computational Linguistics, Proceedings of COLING 2016 System Demonstrations 2016 Dec 11. Association for Computational Linguistics.
- Feng M, Xiang B, Glass MR, Wang L, Zhou B. Applying deep learning to answer selection: A study and an open task. In: IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU); 2015 pp. 813–820.
- Yin W, Kann K, Yu M, Schütze H. Comparative study of cnn and rnn for natural language processing; 2017. arXiv preprint [arXiv:1702.01923](https://arxiv.org/abs/1702.01923).
- Iyyer M, Boyd-Graber J, Claudino L, Socher R, Daumé III H. A neural network for factoid question answering over paragraphs. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), Qatar; 2014. pp. 633–644.
- Yin J, Jiang X, Lu Z, Shang L, Li H, Li X. Neural generative question answering. arXiv preprint [arXiv:1512.01337](https://arxiv.org/abs/1512.01337); 2015.
- Chen D, Fisch A, Weston J, Bordes A. Reading Wikipedia to answer open-domain questions. arXiv preprint [arXiv:1704.00051](https://arxiv.org/abs/1704.00051); 2017.
- Sze V, Chen YH, Yang TJ, Emer JS. Efficient processing of deep neural networks: a tutorial and survey. Proc IEEE. 2017;105(12):2295–329.
- LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86(11):2278–324.
- Elman JL. Finding structure in time. Cognitive Sci. 1990;14(2):179–211.
- Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016 Nov 10.
- Rhee EJ. A deep learning approach for classification of cloud image patches on small datasets. J Inform Commun Conver Eng. 2018;16(3):173–8.
- Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput. 1997;9(8):1735–80.
- Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint [arXiv:1412.3555](https://arxiv.org/abs/1412.3555). 2014.
- Allahyari M, Pouriyeh S, Assefi M, Safaei S, Trippe ED, Gutierrez JB, Kochut K. A brief survey of text mining: Classification, clustering, and extraction techniques. arXiv preprint [arXiv:1707.02919](https://arxiv.org/abs/1707.02919). 2017 Jul 10.
- Suchanek F, Weikum G. Knowledge harvesting in the big-data era. In: Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data 2013 Jun 22 (pp. 933–938).
- Reshmi S, Balakrishnan K. Implementation of an inquisitive chatbot for database supported knowledge bases. sādhanā. 2016 Oct 1;41(10):1173–8.

23. Mizuno J, Tanaka M, Ohtake K, Oh JH, Kloetzer J, Hashimoto C, Torisawa K, WISDOM X, DISAANA, and D-SUMM: Large-scale NLP systems for analyzing textual big data. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations 2016 Dec (pp. 263–267).
24. de Jong M, Zhang K, Roth AM, Rhodes T, Schmucker R, Zhou C, Ferreira S, Cartucho J, Veloso M. Towards a robust interactive and learning social robot. In Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, Sweden; 2018 pp. 883–891.
25. NAOqi developer Guide. http://doc.aldebaran.com/2-5/index_dev_guide.html Accessed on 1 June 2019.
26. Google Speech to Text using Python. <https://pythonspot.com/speech-recognition-using-google-speech-api/> Accessed on 1 June 2019.
27. Budiharto W, Cahyani AD. Behavior-Based Humanoid Robot for Teaching Basic Mathematics, Internetwork Indonesia Journal. Indonesia. 2017;9(1):33–7.
28. García DH, Monje CA, Balaguer C. Knowledge Base Representation for Humanoid Robot Skills, IFAC Proceedings Volumes; 2014 Vol 47(3), pp 3042–3047. Baru sampe sini.
29. Andreas V, Gunawan AA, Budiharto W. Anita: Intelligent Humanoid Robot with Self-Learning Capability Using Indonesian Language. In 2019 4th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS), Tokyo; 2019. pp. 144–147. IEEE.
30. Rajpurkar P, Zhang J, Lopyrev K, Liang P. Squad: 100,000+ questions for machine comprehension of text. arXiv preprint [arXiv:1606.05250](https://arxiv.org/abs/1606.05250). 2016.
31. Pennington J, Socher R, Manning CD. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), Qatar; 2014. pp. 1532–1543.
32. Seo M, Kembhavi A, Farhadi A, Hajishirzi H. Bidirectional attention flow for machine comprehension. arXiv preprint [arXiv:1611.01603](https://arxiv.org/abs/1611.01603); 2016.
33. Sasikumar U, Sindhu L. A survey of natural language question answering system. Int J Comput Appl. 2014;08(15):42–6.
34. NLP–Building a Question-Answering model. <https://towardsdatascience.com/nlp-building-a-question-answering-model-ed0529a68c54> Accessed on 1 June 2019.
35. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805). 2018.
36. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In Advances in neural information processing systems 2017 (pp. 5998–6008).
37. Rajpurkar P, Jia R, Liang P. Know what you don't know: Unanswerable questions for SQuAD. arXiv preprint [arXiv:1806.03822](https://arxiv.org/abs/1806.03822). 2018.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)