Deep Learning-Based Question Answering System for Intelligent Humanoid Robot

Widodo Budiharto*[1], Vincent Andreas[1], and Alexander Agung Santoso Gunawan[2]

[1]Computer Science Department, School of Computer Science, Bina Nusantara University,

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

Email:wbudiharto@binus.edu

Abstract

The development of intelligent Humanoid Robot focuses on question answering systems to be able to interact with people is very rare. In this research, we would like to propose a Humanoid Robot with the self-learning capability for accepting and giving a response from people based on Deep Learning and big data from the internet. 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, where the question from the user will be processed using deep learning, and the result will be compared with knowledge on the system. We proposed our deep learning approach, based on use GRU/LSTM, CNN and BiDAF with big data SQuAD as training dataset. Our experiment indicates that using GRU/LSTM encoder with BiDAF gives higher Exact Match and F1 Score, than CNN with the BiDAF model.

Keywords: Humanoid Robot; NLP; big data

---

1. Introduction

Robot learning is a research field at the intersection of machine learning and robotics. Its studies techniques allowing a robot to acquire novel skills or adapt to its environment through learning algorithms. Perera et al [1] propose a novel approach to enable 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. Roboticist 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].

Our state of the art is to make a question answering system with self-learning in the Humanoid Robot that can talk as a receptionist and answering questions from the dataset using deep learning. Previously we have tried to find a question answering system for Humanoid Robots, but we have not found another journal that discusses it. There are several robots that are similar to Anita, such as Sophia from Hanson Robotics or Asimo from Honda[4] but with a closed question answering system and it is not explained how the system question answering on the Humanoid Robot works.

2. Deep Learning and Knowledge Base System

2.1 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 a great example 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. Feng et al [5] try to do answer selection using some methods. First is by converting question and candidate answers to word vector, then count the cosine similarity of the answer. The highest cosine value means the match between question and answer. The second method is by using an information retrieval baseline, then using cosine similarity to find the highest value. For the deep learning method, they use CNN to handle it. The result shows that a deep learning-based model shows a better result than another method.

Yin W et al [6] do the comparative study of CNN and RNN for NLP. They do some task (Sentiment Classification**,** Relation Classification, Textual Entailment, Answer Selection, Question Relation Match, Path Query Answering, Part-of-Speech Tagging) to test the performances. To get a fair result, they train the models from scratch, using the basic setup, and search optimal hyperparameters for each task and model separately. They found that RNN performs better than CNN in most of the task, except in key phrase recognition task and question-answer match setting.

Iyyer et al [7] do the research about factoid question answering over paragraphs. They use Dependency Tree Recursive Neural Network (DT-RNN) to train the model and use dataset containing dataset from quiz bowl tournaments, do the standardization process and generate features of the dataset. After that, they use Wikipedia page titles to train the model. They test the model in two categories quiz, history, and literature. The test shows that the model does better than the average human player in history question, but the model test results worse on the literature question than humans. From the experiments, they found that DT-RNN is an effective model for question answering, especially in quiz bowl.

Yin J et al [8] use Generative QA (GEN QA) model to generate answers from factoid questions. First, they transform question to representation using bidirectional RNN, then pick the question representation, and interact with Knowledge Base, it can be implemented using Bilinear Model and CNN based Matching Model. The last stage will use RNN to generate an answer. The test result shows that GEN QA based on the CNN model gets the highest score than using GEN QA based on the bilinear model.

Chen et al [9] make unique research, using Wikipedia as the knowledge source to answer the question. They use 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 try to predict the span by using the bilinear term to capture the similarity between paragraphs and questions.

2.2 Knowledge Base System

For the knowledge base system, we can input by directly input the knowledge, by saying the knowledge to the robot, or inserting manual to the knowledge base.

3. Proposed Method

Our research focuses on question answering using Deep Learning approach. The difference between our Humanoid Robot with Pepper [10] the robot is, Pepper covers Human Recognition, Object Recognition and Speech Recognition based on NAOqi software[11]. In this research, we use Google API Speech Recognition using Python language to recognize voice [12]. 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 Robo. The fun aspect is given as well because kids learn best when they are relaxed and focused. They can give a good impact on student learning [13]. 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 [14]. Our previous research is to solve the problem of arithmetic word problems in the subtraction or addition operation given in the Indonesian Language by using various NLP principles and basic self-learning capability [15].

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 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
    say greeting

    while true, do
       listen to user sentence
```

```
if the user asks a question, do
   listen user 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 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 figure 1:

Figure 1. Our model using Deep Learning.

For the training, we will use Stanford Question Answering Dataset (SQuAD) [16] for the machine comprehension dataset. This dataset based on Wikipedia articles with various topics. This dataset contains 87.000 training questions answers (train dataset), and 10.000 development dataset (dev set). The answers' sentences always part of the paragraph article.

In the embedding layer, we convert each word in the text from dataset to word embedding. Word embedding is the representation of the word in the set of the vector, word which similar meaning will have similar representation. We use 100 dimensions of GloVe[17] word embedding.

The next step is the Encoder Layer. The purpose of this step is to make each word in the dataset to know and aware of the previous and next word. We try to use Gated Recurrent Unit (GRU)[18] / Long Short Term Memory (LSTM)[19] encoder and CNN 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 answer based on hidden vector for dataset and hidden vector for a question. We must use the Attention layer to find the answer. We use BiDAF (Bidirectional Attention Flow) [20] as shown in figure 2.

Figure 2. Bidirectional Attention Flow[20].

The first step to use the BiDAF Attention Layer is computing similarity matrix. S ∈ R N×M, which contains a similarity score S_{ij} for each pair (c_i , q_j) of context and question hidden states. S_{ij} = wT sim[$_{ci}$; q_j ; $c_i \circ q_j$] ∈ R. Here, $c_i \circ q_j$ is an elementwise product and wsim ∈ R 6h is a weight vector.

Next, we perform Context-to-Question (C2Q) Attention. We take the row-wise softmax of S to obtain attention distributions α_i , which we use to take weighted sums of the question hidden states q j , yielding C2Q attention outputs a_i. The next step is Question-to-Context(Q2C) Attention. For each context location i ∈ {1, . . . , N}, we take the max of the corresponding row of the similarity matrix, m_i = max j S_{ij} ∈ R. Then we take the softmax over the resulting vector m ∈ R N — this gives us an attention distribution β ∈ 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.

$$\beta = softmax(m) \in R^N$$

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

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

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

The final layer is a softmax output layer that helps us decide the start and the end index for the answer span. We combine the context of hidden states and the attention vector from the previous layer 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* [21][22].

4. Experimental Results and Discussion

For the experiments, we try to test using RNN encoder and CNN encoder. We use 150 hidden encoders, 150 hidden models, 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 two to three days.

During the training, we check the Exact Match Score (EM Score) and F1 score using the development dataset, and after the model finished, we test them using the test dataset, which we get from 10% of the training dataset. The result shows below.

Table 1. Result of our experiment

Method	Dev (em /f1)	Test (em / f1)
RNN Encoder	51.69 / 67.24	68.87 / 82.43
CNN Encoder	46.35 / 61.34	53.99 / 69.55

For models using the RNN Encoder, we get the optimum model at 93.000th iteration, while the model using the CNN Encoder, we get the optimal result at 43.000th iteration. From two different approach, we find that using RNN Encoder give better EM and F1 score result. 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.

5. Conclusion

Our model is successfully able to obtain knowledge using big data technology and response question from the user using deep learning. From our experiment using RNN and CNN as an

encoder layer, we found that with RNN and BiDAF attention layer, we get high EM and F1 scores, so the model can be used to handle question answering between Humanoid Robot and human. The RNN encoder will give a higher EM / F1 score than using the CNN encoder.

For future development, we will use the database to save knowledge, so the knowledge can store more data and manage easily. We also will improve our algorithm to make better results to find the answer and improve the ability of the Humanoid Robot to handle unanswerable questions using the SQuAD 2.0 Dataset [23], and we will crawl the website for the knowledge base of the Humanoid Robot in the future.

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no competing interests.

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.

Authors' contributions

Both authors read and approved the final manuscript.

Acknowledgment

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.

Author's Information

[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.

Corresponding author

Correspondence to Widodo Budiharto, email: wbudiharto@binus.edu.

Abbreviations

NLP: Natural Language Processing

GRU: Gated Recurrent Unit

LSTM: Long Short Term Memory

CNN: Convolution Neural Network

BiDAF: Bidirectional Attention Flow

EM: Exact Match

QA: Question Answering

Q2C: Question-to-Context

SQuAD: Stanford Question Answering Dataset

References

[1] Perera V and Veloso M. Learning to Understand Questions on the Task History of a Service Robot. IEEE International Symposium on Robot and Human Interactive Communication (RO-

MAN), Portugal. 2017; p. 304-309.

[2] Kwon D. Self-Taught Robots, Scientific American. 2018; p. 26-31.

[3] Billard A et al. Robot programming by demonstration. In Siciliano B, Khatib O (eds) Handbook of Robotics, Springer, Secaucus, NJ, USA. 2008; p. 1371–1394.

[4] Sakagami et al. The intelligent ASIMO: System overview and integration. Intl. Conference on Intelligent Robots and Systems. Switzerland. 2002; p. 2478 – 2483.

[5] Feng M et al. Applying Deep Learning to Answer Selection: A Study and An Open Task. 2015; arXiv preprint arXiv:1508.01585v2 [cs.CL].

[6] Yin W et al. Comparative Study of CNN and RNN for Natural Language Processing. 2017; arXiv preprint arXiv:1702.01923v1.

[7] Iyyer M et al. A Neural Network for Factoid Question Answering over Paragraphs. Conference on Empirical Methods in Natural Language Processing (EMNLP), Qatar, 2014; p 633–644.

[8] Yin J et al. Neural Generative Question Answering. 2016; arXiv preprint arXiv:1512.01337v4 [cs.CL].

[9] Chen D et al. Reading Wikipedia to Answer Open-Domain Questions. 2017; arXiv preprint arXiv:1704.00051v2 [cs.CL]

[10] Jong MD et al. Towards a Robust Interactive and Learning Social Robot. AAMAS, Sweden. 2018; p 883 – 891.

[11] NAOqi developer Guide http://doc.aldebaran.com/2-5/index_dev_guide.html Accessed 1 June 2019.

[12] Google Speech to Text using Python https://pythonspot.com/speech-recognition-using-google-speech-api/ Accessed 1 June 2019.

[13] Budiharto W, Cahyani AD. Behavior-Based Humanoid Robot for Teaching Basic

Mathematics, Internetwork Indonesia Journal. 2017; 9(1), p 33-37.

[14] García DH, Monje CA, Balaguer C. Knowledge Base Representation for Humanoid Robot Skills, IFAC Proceedings Volumes. 2014; 47(3), p 3042-3047.

[15] Andreas V, Gunawan AA, Budiharto W. Anita: Intelligent Humanoid Robot with Self-Learning Capability using Indonesian Language, ACIRS 2019, Tokyo. 2019; p.144-147.

[16] Rajpurkar P et al (2016) SQuAD: 100,000+ Questions for Machine Comprehension of Text, arXiv preprint arXiv:1606.05250v3 [cs.CL].

[17] Pennington J et al (2014). Glove: Global vectors for word representation. In Empirical Methods in Natural Language Processing (EMNLP). 2014; p. 1532–1543.

[18] Chung et al. Empirical evaluation of gated recurrent neural network on sequence modeling. 2014; CoRR, abs/1412.3555

[19] Hochreiter S, Schmidhuber J. Long Short-Term Memory, Neural Computation,1997; 9(8), p.1735–1780.

[20] Seo M et al. Bi-Directional Attention Flow for Machine Comprehension, ICLR, France. 2017.

[21] Sasikumar U, Sindhu L. A Survey of Natural Language Question Answering System, International Journal of Computer Application, 2014;08(15). p 42-46

[22] NLP – Building a Question-Answering model https://towardsdatascience.com/nlp-building-a-question-answering-model-ed0529a68c54 Accessed 1 June 2019

[23] Rajpurkar et al. Know What You Don’t Know: Unanswerable Questions for SQuAD. 2018; preprint arXiv arXiv:1806.03822v1 [cs.CL].